

Part 2

Species Interactions



Introduction

The activity of any organism changes the environment in which it lives. It may alter conditions, as when the transpiration of a tree cools the atmosphere, or it may add or subtract resources from the environment that might have been available to other organisms, as when that tree shades the plants beneath it. In addition, though, organisms interact when individuals enter into the lives of others. In the following chapters (8–15) we consider the variety of these interactions between individuals of different species. We distinguish five main categories: competition, predation, parasitism, mutualism and detritivory, although like most biological categories, these five are not perfect pigeon-holes.

In very broad terms, ‘competition’ is an interaction in which one organism consumes a resource that would have been available to, and might have been consumed by, another. One organism deprives another, and, as a consequence, the other organism grows more slowly, leaves fewer progeny or is at greater risk of death. The act of deprivation can occur between two members of the same species or between individuals of different species. We have already examined *intraspecific* competition in Chapter 5. We turn to *interspecific* competition in Chapter 8.

Chapters 9 and 10 deal with various aspects of ‘predation’, though we have defined predation broadly. We have combined those situations in which one organism eats another and kills it (such as an owl preying on mice), and those in which the consumer takes only part of its prey, which may then regrow to provide another bite another day (grazing). We have also combined herbivory (animals eating plants) and carnivory (animals eating animals). In Chapter 9 we examine the nature of predation, i.e. what happens to the predator and what happens to the prey, paying particular attention to herbivory because of the subtleties that

characterize the response of a plant to attack. We also discuss the behavior of predators. Then, in Chapter 10, we examine the ‘consequences of consumption’ in terms of the dynamics of predator and prey populations. This is the part of ecology that has the most obvious relevance to those concerned with the management of natural resources: the efficiency of harvesting (whether of fish, whales, grasslands or prairies) and the biological and chemical control of pests and weeds – themes that we take up in Chapter 15.

Most of the processes in this section involve genuine *interactions* between organisms of different species. However, when dead organisms (or dead parts of organisms) are consumed – decomposition and detritivory – the affair is far more one-sided. None the less, as we describe in Chapter 11, these processes themselves incorporate competition, parasitism, predation and mutualism: microcosms of all the major ecological processes (except photosynthesis).

Chapter 12, ‘Parasitism and Disease’, deals with a subject that in the past was often neglected by ecologists – and by ecology texts. Yet more than half of all species are parasites, and recent years have seen much of that past neglect rectified. Parasitism itself has blurred edges, particularly where it merges into predation. But whereas a predator usually takes all or part of many individual prey, a parasite normally takes its resources from one or a very few hosts, and (like many grazing predators) it rarely kills its hosts immediately, if at all.

Whereas the earlier chapters of this section deal largely with conflict between species, Chapter 13 is concerned with mutualistic interactions, in which both organisms experience a net benefit. None the less, as we shall see, conflict often lies at the heart of mutualistic interactions too: each participant exploiting the other, such that the *net* benefit arises only because, overall, gains exceed losses. Like parasitism, the ecology of mutualism

has often been neglected. Again, though, this neglect has been unwarranted: the greater part of the world's biomass is composed of mutualists.

Ecologists have often summarized interactions between organisms by a simple code that represents each one of the pair of interacting organisms by a '+', a '-' or a '0', depending on how it is affected by the interaction. Thus, a predator-prey (including a herbivore-plant) interaction, in which the predator benefits and the prey is harmed, is denoted by + -, and a parasite-host interaction is also clearly + -. Another straightforward case is mutualism, which, overall, is obviously + +; whereas if organisms do not interact at all, we can denote this by 0 0 (sometimes called 'neutralism'). Detritivory must be denoted by + 0, since the detritivore itself benefits, while its food (dead already) is unaffected. The general term applied to + 0 interactions is 'commensalism', but paradoxically this term is not usually used for detritivores. Instead, it is reserved for cases, allied to parasitism, in which one organism (the 'host') provides resources or a home for another organism, but in which the host itself suffers no tangible ill effects. Competition is usually described as a - - interaction, but it is often impossible to establish that both organisms are harmed. Such asymmetric interactions may then approximate to a - 0 classification, generally referred to as 'amensalism'. True cases

of amensalism may occur when one organism produces its ill effect (for instance a toxin) whether or not the potentially affected organism is present.

Although the earlier chapters in this section deal with these various interactions largely in isolation, members of a population are subject simultaneously to many such interactions, often of all conceivable types. Thus, the abundance of a population is determined by this range of interactions (and indeed environmental conditions and the availability of resources) all acting in concert. Attempts to understand variations in abundance therefore demand an equally wide ranging perspective. We adopt this approach in Chapter 14.

Finally in this section, we discuss in Chapter 15 applications of the principles elaborated in the preceding chapters. Our focus is on pest control and the management of natural resources. With the former, the pest species is either a competitor or a predator of desirable species (for example food crops), and we are either predators of the pest ourselves or we manipulate its natural predators to our advantage (biological control). With the latter, again, we are predators of a living, natural resource (harvestable trees in a forest, fish in the sea), but the challenge for us is to establish a stable and sustainable relationship with the prey, guaranteeing further valuable harvests for generations to come.

Chapter 8

Interspecific Competition



8.1 Introduction

The essence of interspecific competition is that individuals of one species suffer a reduction in fecundity, growth or survivorship as a result of resource exploitation or interference by individuals of another species. This competition is likely to affect the population dynamics of the competing species, and the dynamics, in their turn, can influence the species' distributions and their evolution. Of course, evolution, in *its* turn, can influence the species' distributions and dynamics. Here, we concentrate on the effects of competition on populations of species, whilst Chapter 19 examines the role of interspecific competition (along with predation and parasitism) in shaping the structure of ecological communities. There are several themes introduced in this chapter that are taken up and discussed more fully in Chapter 20. The two chapters should be read together for a full coverage of interspecific competition.

8.2 Some examples of interspecific competition

a diversity of examples of competition . . .

There have been many studies of interspecific competition between species of all kinds. We have chosen six initially, to illustrate a number of important ideas.

8.2.1 Competition between salmonid fishes

. . . between salmonid fishes, . . .

Salvelinus malma (Dolly Varden charr) and *S. leucomaenis* (white-spotted charr) are morphologically similar and closely related fishes in the family Salmonidae. The two species are found together in many streams on Hokkaido Island in Japan, but Dolly Varden are distributed at higher altitudes (further

upstream) than white-spotted charr, with a zone of overlap at intermediate altitudes. In streams where one species happens to be absent, the other expands its range, indicating that the distributions may be maintained by competition (i.e. each species suffers, and is thus excluded from certain sites, in the presence of the other species). Water temperature, an abiotic factor with profound consequences for fish ecology (discussed already in Section 2.4.4), increases downstream.

By means of experiments in artificial streams, Taniguchi and Nakano (2000) showed that when either species was tested alone, higher temperatures led to increased aggression. But this effect was reversed for Dolly Varden when in the presence of white-spotted charr (Figure 8.1a). Reflecting this, at the higher temperature, Dolly Varden were suppressed from obtaining favorable foraging positions when white-spotted charr were present, and they suffered lower growth rates (Figure 8.1b, c) and a lower probability of survival.

Thus, the experiments lend support to the idea that Dolly Varden and white-spotted charr compete: one species, at least, suffers directly from the presence of the other. They coexist in the same river, but on a finer scale their distributions overlap very little. Specifically, the white-spotted charr appear to outcompete and exclude Dolly Varden from downstream locations in the latter's range. The reason for the upper boundary of white-spotted charr remains unknown as they did not suffer from the presence of Dolly Varden at the lower temperature.

8.2.2 Competition between barnacles

The second study concerns two species of barnacle in Scotland: *Chthamalus stellatus* and *Balanus balanoides* (Figure 8.2) (Connell, 1961). These are frequently

. . . between barnacles, . . .

found together on the same Atlantic rocky shores of northwest

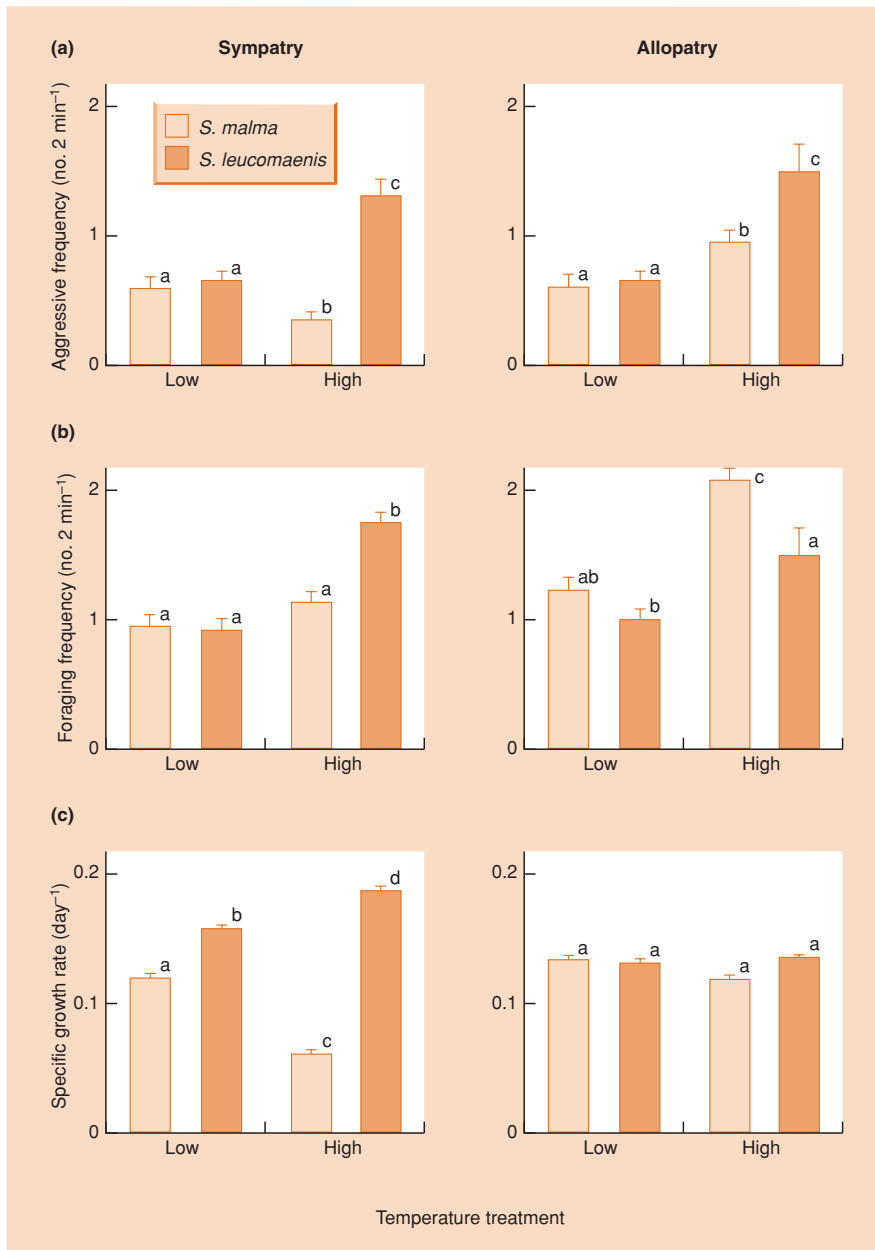


Figure 8.1 (a) Frequency of aggressive encounters initiated by individuals of each fish species during a 72-day experiment in artificial stream channels with two replicates each of 50 Dolly Varden (*Salvelinus malma*) or 50 white-spotted charr (*S. leucomaenis*) alone (allopatry) or 25 of each species together (sympatry). (b) Foraging frequency. (c) Specific growth rate in length. Different letters indicate that the means are significantly different from each other. (From Taniguchi & Nakano, 2000.)

Europe. However, adult *Chthamalus* generally occur in an intertidal zone that is higher up the shore than that of adult *Balanus*, even though young *Chthamalus* settle in considerable numbers in the *Balanus* zone. In an attempt to understand this zonation, Connell monitored the survival of young *Chthamalus* in the *Balanus* zone. He took successive censuses of mapped individuals over the period of 1 year and, most importantly, he ensured at some sites that young *Chthamalus* that settled in the *Balanus* zone were kept free from contact with *Balanus*. In contrast with the normal pattern, such individuals survived well, irrespective of the intertidal

level. Thus, it seemed that the usual cause of mortality in young *Chthamalus* was not the increased submergence times of the lower zones, but competition from *Balanus* in those zones. Direct observation confirmed that *Balanus* smothered, undercut or crushed *Chthamalus*, and the greatest *Chthamalus* mortality occurred during the seasons of most rapid *Balanus* growth. Moreover, the few *Chthamalus* individuals that survived 1 year of *Balanus* crowding were much smaller than uncrowded ones, showing, since smaller barnacles produce fewer offspring, that interspecific competition was also reducing fecundity.

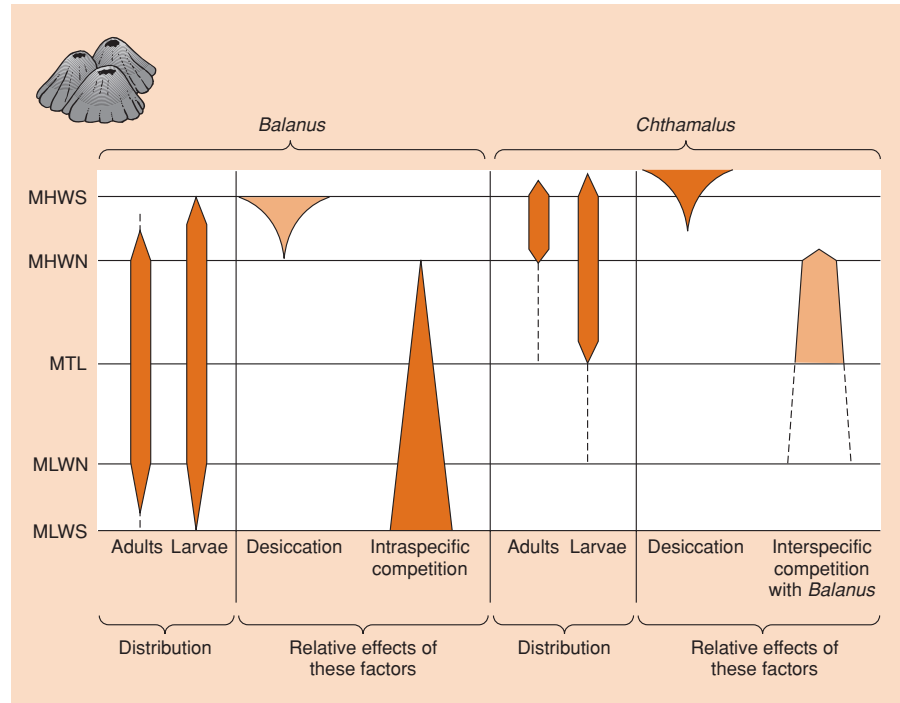


Figure 8.2 The intertidal distribution of adults and newly settled larvae of *Balanus balanoides* and *Chthamalus stellatus*, with a diagrammatic representation of the relative effects of desiccation and competition. Zones are indicated to the left: from MHWS (mean high water, spring) down to MLWS (mean low water, spring); MTL, mean tide level; N, neap. (After Connell, 1961.)

Thus, *Balanus* and *Chthamalus* compete. They coexist on the same shore but, like the fish in the previous section, on a finer scale their distributions overlap very little. *Balanus* outcompetes and excludes *Chthamalus* from the lower zones; but *Chthamalus* can survive in the upper zones where *Balanus*, because of its comparative sensitivity to desiccation, cannot.

8.2.3 Competition between bedstraws (*Galium* spp.)

... between bedstraws, ...

A. G. Tansley, one of the greatest of the ‘founding fathers’ of plant ecology, studied competition between two species of bedstraw (Tansley, 1917). *Galium hercynicum* is a species which grows naturally in Great Britain at acidic sites, whilst *G. pumilum* is confined to more calcareous soils. Tansley found in experiments that as long as he grew them alone, both species would thrive on both the acidic soil from a *G. hercynicum* site and the calcareous soil from a *G. pumilum* site. Yet, if the species were grown together, only *G. hercynicum* grew successfully in the acidic soil and only *G. pumilum* grew successfully in the calcareous soil. It seems, therefore, that when they grow together the species compete, and that one species wins, whilst the other loses so badly that it is competitively excluded from the site. The outcome depends on the habitat in which the competition occurs.

8.2.4 Competition between *Paramecium* species

The fourth example comes from the classic work of the great Russian ecologist G. F. Gause, who studied competition in laboratory experiments using three species of the protozoan *Paramecium* (Gause, 1934, 1935). All three species grew well alone, reaching stable carrying capacities in tubes of liquid medium. There, *Paramecium* consumed bacteria or yeast cells, which themselves lived on regularly replenished oatmeal (Figure 8.3a).

When Gause grew *P. aurelia* and *P. caudatum* together, *P. caudatum* always declined to the point of extinction, leaving *P. aurelia* as the victor (Figure 8.3b). *P. caudatum* would not normally have starved to death as quickly as it did, but Gause’s experimental procedure involved the daily removal of 10% of the culture and animals. Thus, *P. aurelia* was successful in competition because near the point where its population size leveled off, it was still increasing by 10% per day (and able to counteract the enforced mortality), whilst *P. caudatum* was only increasing by 1.5% per day (Williamson, 1972).

By contrast, when *P. caudatum* and *P. bursaria* were grown together, neither species suffered a decline to the point of extinction – they coexisted. But, their stable densities were much lower than when grown alone (Figure 8.3c), indicating that they were in competition with one another (i.e. they ‘suffered’). A closer

... between *Paramecium* species, ...

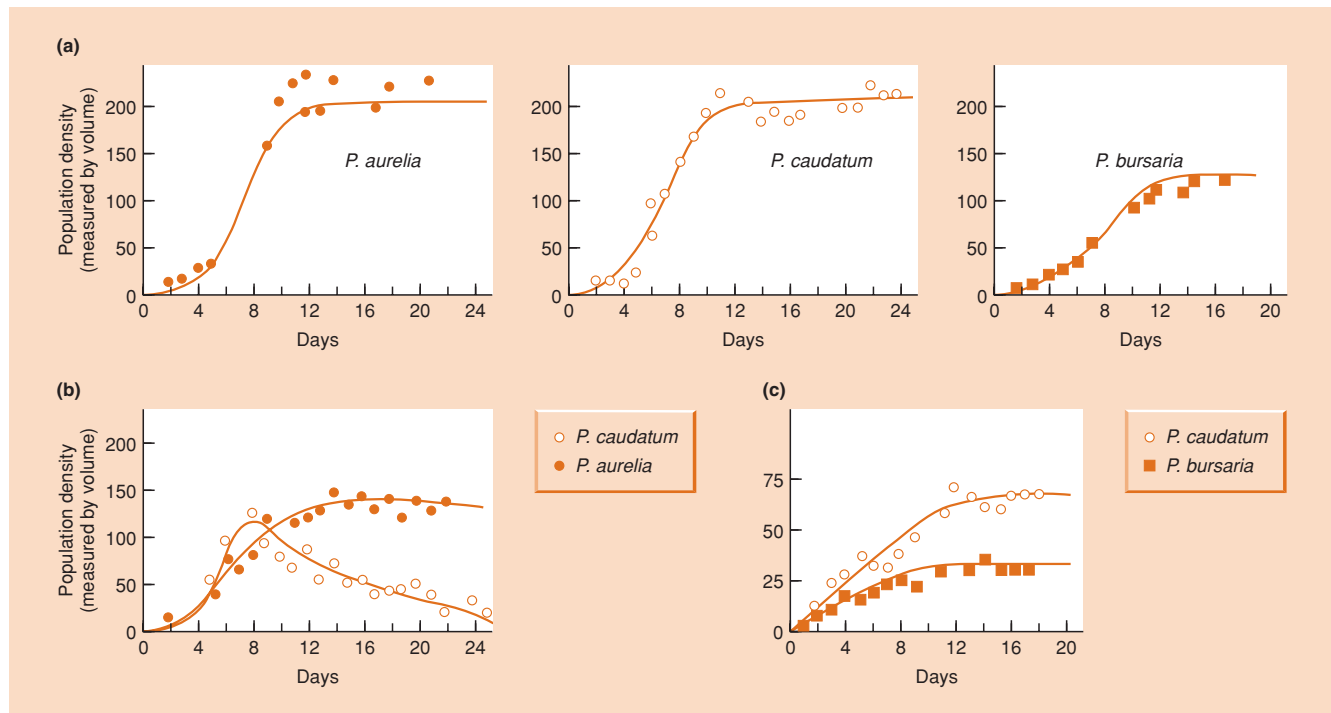


Figure 8.3 Competition in *Paramecium*. (a) *P. aurelia*, *P. caudatum* and *P. bursaria* all establish populations when grown alone in culture medium. (b) When grown together, *P. aurelia* drives *P. caudatum* towards extinction. (c) When grown together, *P. caudatum* and *P. bursaria* coexist, although at lower densities than when alone. (After Clapham, 1973; from Gause, 1934.)

look, however, revealed that although they lived together in the same tubes, they were, like Taniguchi and Nakano's fish and Connell's barnacles, spatially separated. *P. caudatum* tended to live and feed on the bacteria suspended in the medium, whilst *P. bursaria* was concentrated on the yeast cells at the bottom of the tubes.

8.2.5 Coexistence amongst birds

... among birds ... Ornithologists are well aware that closely related species of birds often coexist in the same habitat. For example, five *Parus* species occur together in English broad-leaved woodlands: the blue tit (*P. caeruleus*), the great tit (*P. major*), the marsh tit (*P. palustris*), the willow tit (*P. montanus*) and the coal tit (*P. ater*). All have short beaks and hunt for food chiefly on leaves and twigs, but at times on the ground; all eat insects throughout the year, and also seeds in winter; and all nest in holes, normally in trees. However, the closer we look at the details of the ecology of such coexisting species, the more likely we will find ecological differences – for example, in precisely where within the trees they feed, in the size of their insect prey and the hardness of the seeds they take. Despite their similarities, we may be tempted to conclude that the tit species compete but coexist by eating slightly different resources

in slightly different ways. However, a scientifically rigorous approach to determine the current role of competition requires the removal of one or more of the competing species and monitoring the responses of those that remain. Martin and Martin (2001) did just this in a study of two very similar species: the orange-crowned warbler (*Vermivora celata*) and virginia's warbler (*V. virginiae*) whose breeding territories overlap in central Arizona. On plots where one of the two species had been removed, the remaining orange-crowned or virginia's warblers fledged between 78 and 129% more young per nest, respectively. The improved performance was due to improved access to preferred nest sites and consequent decreased losses of nestlings to predators. In the case of virginia's warblers, but not orange-crowned warblers, feeding rate also increased in plots from which the other species was removed (Figure 8.4).

8.2.6 Competition between diatoms

The final example is from a laboratory investigation of two species of freshwater diatom: *Asterionella formosa* and *Synedra ulna* (Tilman *et al.*, 1981). Both these algal species require silicate in the construction of their cell walls. The investigation was

... and between diatoms

Figure 8.4 (right) Percentage difference in feeding rates (mean \pm SE) at orange-crowned warbler and virginia's warbler nests on plots where the other species had been experimentally removed. Feeding rates (visits per hour to the nest with food) were measured during incubation (inc) (rates of male feeding of incubating females on the nest) and during the nestling period (nstl) (nestling feeding rates by both parents combined). P values are from t -tests of the hypothesis that each species fed at higher rates on plots from which the other had been removed. This hypothesis was supported for virginia's warblers but not orange-crowned warblers. (After Martin & Martin, 2001.)

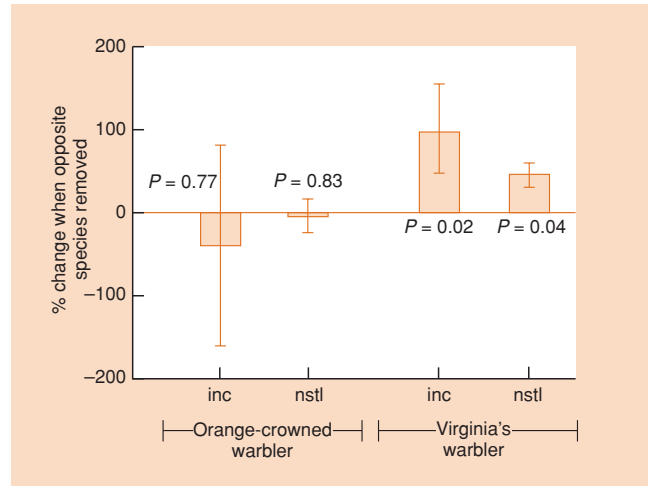
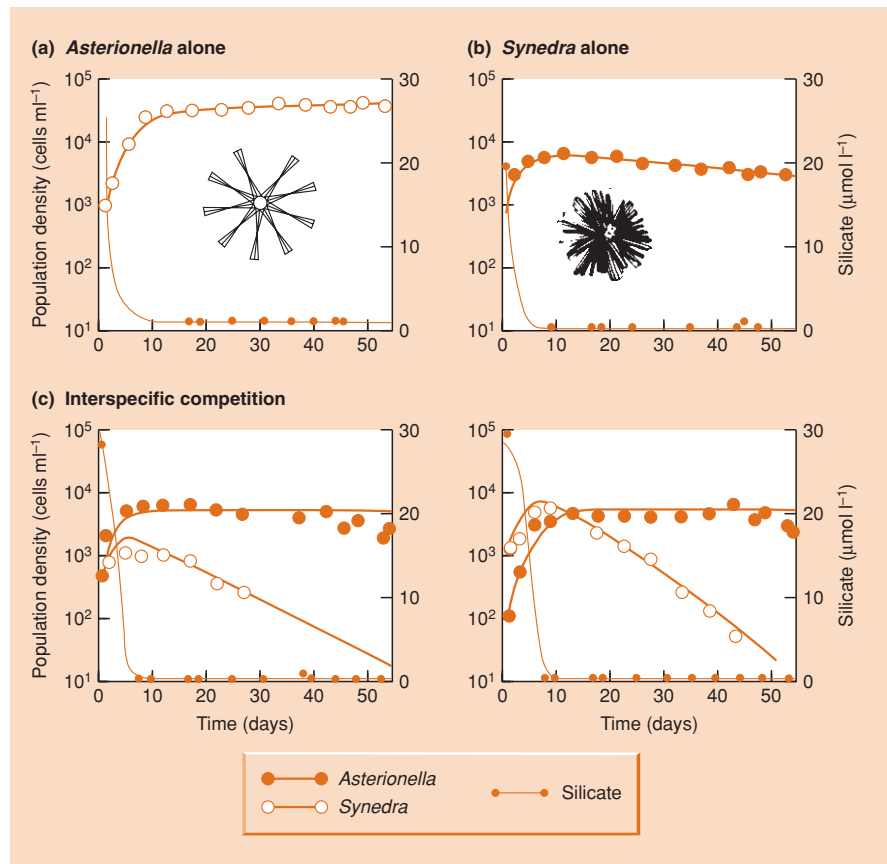


Figure 8.5 Competition between diatoms. (a) *Asterionella formosa*, when grown alone in a culture flask, establishes a stable population and maintains a resource, silicate, at a constant low level. (b) When *Synedra ulna* is grown alone it does the same, but maintains silicate at an even lower level. (c) When grown together, in two replicates, *Synedra* drives *Asterionella* to extinction. (After Tilman *et al.*, 1981.)



unusual because at the same time as population densities were being monitored, the impact of the species on their limiting resource (silicate) was being recorded. When either species was cultured alone in a liquid medium to which resources were continuously being added, it reached a stable carrying capacity whilst maintaining the silicate at a constant low concentration

(Figure 8.5a, b). However, in exploiting this resource, *Synedra* reduced the silicate concentration to a lower level than did *Asterionella*. Hence, when the two species were grown together, *Synedra* maintained the concentration at a level that was too low for the survival and reproduction of *Asterionella*. *Synedra* therefore competitively excluded *Asterionella* from mixed cultures (Figure 8.5c).

8.3 Assessment: some general features of interspecific competition

8.3.1 Unraveling ecological and evolutionary aspects of competition

These examples show that individuals of different species can compete. This is hardly surprising. The field experiments with barnacles and warblers also show that different species do compete in nature (i.e. there was a measurable interspecific reduction in abundance and/or fecundity and/or survivorship). It seems, moreover, that competing species may either exclude one another from particular habitats so that they do not coexist (as with the bedstraws, the diatoms and the first pair of *Paramecium* species), or may coexist, perhaps by utilizing the habitat in slightly different ways (e.g. the barnacles and the second pair of *Paramecium* species).

But what about the story of the coexisting tits? Certainly the five bird species coexist and utilize the habitat in slightly different ways. But does this have anything to do with competition? It may do. It may be that the five species of tit coexist as a result of evolutionary responses to interspecific competition. This requires some further explanation. When two species compete, individuals of one or both species may suffer reductions in fecundity and/or survivorship, as we have seen. The fittest individuals of each species may then be those that (relatively speaking) escape competition because they utilize the habitat in ways that differ most from those adopted by individuals of the other species. Natural selection will then favor such individuals, and eventually the population may consist entirely of them. The two species will evolve to become more different from one another than they were previously; they will compete less, and thus will be more likely to coexist.

coexisting competitors or the 'ghost of competition past'? . . .

The trouble with this as an explanation for the tit story is that there is no proof. We need to beware, in Connell's (1980) phrase, of uncritically invoking the 'ghost of competition past'. We cannot go back in time to

check whether the species ever competed more than they do now. A plausible alternative interpretation is that the species have, in the course of their evolution, responded to natural selection in different but entirely independent ways. They are distinct species, and they have distinctive features. But they do not compete now, nor have they ever competed; they simply happen to be different. If all this were true, then the coexistence of the tits would have nothing to do with competition. Alternatively again, it may be that competition in the past eliminated a number of other species, leaving behind only those that are different in their utilization of the habitat: we can still see the hand of the ghost of competition past, but acting as an ecological force (eliminating species) rather than an evolutionary one (changing them).

The tit story, therefore, and the difficulties with it, illustrate two important general points. The first is that we must pay careful, and separate, attention to both the ecological and the evolutionary effects of interspecific competition. The ecological effects are, broadly, that species may be eliminated from a habitat by competition from individuals of other species; or, if competing species coexist, that individuals of at least one of them suffer reductions in survival and/or fecundity. The evolutionary effects appear to be that species differ more from one another than they would otherwise do, and hence compete less (but see Section 8.9).

The second point, though, is that there are profound difficulties in invoking competition as an explanation for observed patterns, and especially in invoking it as an evolutionary explanation. An experimental manipulation (for instance, the removal of one or more species) can, as we have seen with the warblers, indicate the presence of current competition if it leads to an increase in the fecundity or survival or abundance of the remaining species. But negative results would be equally compatible with the past elimination of species by competition, the evolutionary avoidance of competition in the past, and the independent evolution of noncompeting species. In fact, for many sets of data, there are no easy or agreed methods of distinguishing between these explanations (see Chapter 19). Thus, in the remainder of this chapter (and in Chapter 19) when examining the ecological and, especially, the evolutionary effects of competition, we will need to be more than usually cautious.

. . . or simply evolution?

8.3.2 Exploitation and interference competition and allelopathy

For now, though, what other general features emerge from our examples? As with intraspecific competition, a basic distinction can be made between interference and exploitation competition (although elements of both may be found in a single interaction) (see Section 5.1.1). With exploitation, individuals interact with each other indirectly, responding to a resource level that has been depressed by the activity of competitors. The diatom work provides a clear example of this. By contrast, Connell's barnacles provide an equally clear example of interference competition. *Balanus*, in particular, directly and physically interfered with the occupation by *Chthamalus* of limited space on the rocky substratum.

interference and exploitation

Interference, on the other hand, is not always as direct as this. Amongst plants, it has often been claimed that interference occurs through the production and release into the environment of chemicals that are toxic to other species but not to the producer (known as allelopathy). There is no doubt that chemicals with such

allelopathy

properties can be extracted from plants, but establishing a role for them in nature or that they have evolved *because of* their allelopathic effects, has proved difficult. For example, extracts from more than 100 common agricultural weeds have been reported to have allelopathic potential against crop species (Foy & Inderjit, 2001), but the studies generally involved unnatural laboratory bioassays rather than realistic field experiments. In a similar manner, Vandermeest *et al.* (2002) showed in the laboratory that an extract from American chestnut leaves (*Castanea dentata*) suppressed germination of the shrub rosebay rhododendron (*Rhododendron maximum*). The American chestnut was the most common overstorey tree in the USA's eastern deciduous forest until ravaged by chestnut blight (*Cryphonectria parasitica*). Vandermeest *et al.* concluded that the expansion of rhododendron thickets throughout the 20th century may have been due as much to the cessation of the chestnut's allelopathic influence as to the more commonly cited invasion of canopy openings following blight, heavy logging and fire. However, their hypothesis cannot be tested. Amongst competing tadpole species, too, water-borne inhibitory products have been implicated as a means of interference (most notably, perhaps, an alga produced in the feces of the common frog, *Rana temporaria*, inhibiting the natterjack toad, *Bufo calamita* (Beebee, 1991; Griffiths *et al.*, 1993)), but here again their importance in nature is unclear (Petranka, 1989). Of course, the production by fungi and bacteria of allelopathic chemicals that inhibit the growth of potentially competing microorganisms is widely recognized – and exploited in the selection and production of antibiotics.

8.3.3 Symmetric and asymmetric competition

interspecific competition is frequently highly asymmetric

Interspecific competition (like intraspecific competition) is frequently highly asymmetric – the consequences are often not the same for both species. For instance, with Connell's barnacles,

Balanus excluded *Chthamalus* from their zone of potential overlap, but any effect of *Chthamalus* on *Balanus* was negligible: *Balanus* was limited by its own sensitivity to desiccation. An analogous situation is provided by two species of cattail (reedmace) in ponds in Michigan; *Typha latifolia* occurs mostly in shallower water whilst *T. angustifolia* occurs in deeper water. When grown together (in sympatry) in artificial ponds, the two species mirror their natural distributions, with *T. latifolia* mainly occupying depth zones from 0 to 60 cm below the water surface and *T. angustifolia* mainly from 60 to 90 cm (Grace & Wetzel, 1998). When grown on its own (allopatry), the depth distribution of *T. angustifolia* shifts markedly towards shallower depths. In contrast, *T. latifolia* shows only a minor shift towards greater depth in the absence of interspecific competition.

On a broader front, it seems that highly asymmetric cases of interspecific competition (where one species is little affected)

generally outnumber symmetric cases (e.g. Keddy & Shipley, 1989). The more fundamental point, however, is that there is a continuum linking the perfectly symmetric competitive cases to strongly asymmetric ones. Asymmetric competition results from the differential ability of species to occupy higher positions in a competitive hierarchy. In plants, for example, this may result from height differences, with one species able to completely over-top another and preempt access to light (Freckleton & Watkinson, 2001). In a similar vein, Dezfuli *et al.* (2002) have argued that asymmetric competition might be expected between parasite species that occupy sequential positions in the gut of their host, with a stomach parasite reducing resources and adversely influencing an intestinal parasite further downstream, but not vice versa. Asymmetric competition is especially likely where there is a very large difference in the size of competing species. Reciprocal exclusion experiments have shown that grazing ungulates (domestic sheep and Spanish ibex *Capra pyrenaica*) reduce the abundance of the herbivorous beetle *Timarcha lugens* in Spanish scrubland by exploitation competition (and partly by incidental predation). However, there was no effect of beetle exclusion on ungulate performance (Gomez & Gonzalez-Megias, 2002).

8.3.4 Competition for one resource may influence competition for another

Finally, it is worth noting that competition for one resource often affects the ability of an organism to exploit another resource. For example, Buss (1979) showed that in interactions between species of bryozoa (colonial, modular animals), there appears to be an interdependence between competition for space and for food. When a colony of one species contacts a colony of another species, it interferes with the self-generated feeding currents upon which bryozoans rely (competition for space affects feeding). But a colony short of food will, in turn, have a greatly reduced ability to compete for space (by overgrowth).

Comparable examples are found amongst rooted plants. If one species

root and shoot competition

invades the canopy of another and deprives it of light, the suppressed species will suffer directly from the reduction in light energy that it obtains, but this will also reduce its rate of root growth, and it will therefore be less able to exploit the supply of water and nutrients in the soil. This in turn will reduce its rate of shoot and leaf growth. Thus, when plant species compete, repercussions flow backwards and forwards between roots and shoots (Wilson, 1988a). A number of workers have attempted to separate the effects of canopy and root competition by an experimental design in which two species are grown: (i) alone; (ii) together; (iii) in the same soil, but with their canopies separated; and (iv) in separate soil with their canopies intermingling. One example is a study of

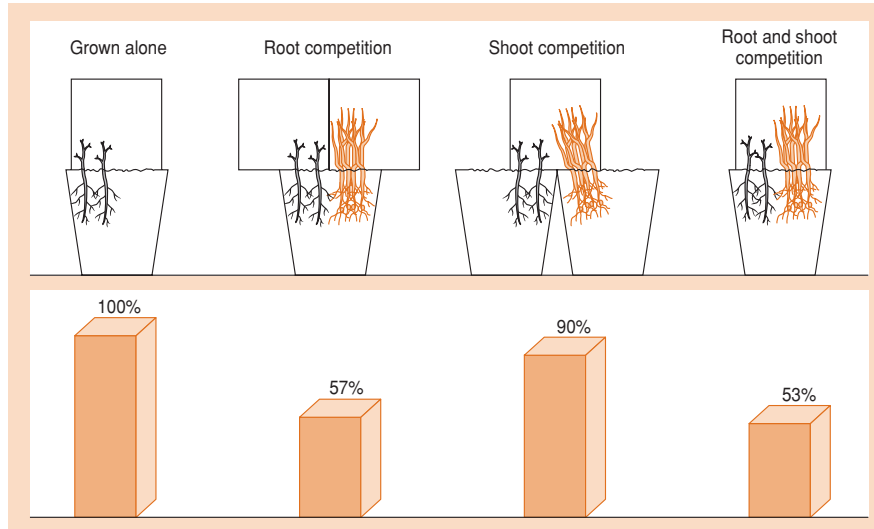


Figure 8.6 Root and shoot competition between maize and pea plants. Above are the experimental plants used, below are the dry weights of pea plants after 46 days as a percentage of those achieved when grown alone. (Data from Semere & Froud-Williams, 2001.)

maize (*Zea mays*) and pea plants (*Pisum sativum*) (Semere & Froud-Williams, 2001). In full competition, with roots and shoots intermingling, the biomass production of maize and peas respectively (dry matter per plant, 46 days after sowing) was reduced to 59 and 53% of the ‘control’ biomass when the species were grown alone. When only the roots intermingled, pea plant biomass production was still reduced to 57% of the control value, but when just the shoots intermingled, biomass production was only reduced to 90% of the control (Figure 8.6). These results indicate, therefore, that soil resources (mineral nutrients and water) were more limiting than light, a common finding in the literature (Snaydon, 1996). They also support the idea of root and shoot competition combining to generate an overall effect, in that the overall reduction in plant biomass (to 53%) was close to the product of the root-only and shoot-only reductions (90% of 57% is 51.3%).

8.4 Competitive exclusion or coexistence?

The results of experiments such as those described here highlight a critical question in the study of the ecological effects of interspecific competition: what are the general conditions that permit the coexistence of competitors, and what circumstances lead to competitive exclusion? Mathematical models have provided important insights into this question.

8.4.1 A logistic model of interspecific competition

The ‘Lotka–Volterra’ model of interspecific competition (Volterra, 1926; Lotka, 1932) is an extension of the logistic equation described

in Section 5.9. As such, it incorporates all of the logistic’s shortcomings, but a useful model can none the less be constructed, shedding light on the factors that determine the outcome of a competitive interaction.

The logistic equation:

$$\frac{dN}{dt} = rN \frac{(K - N)}{K} \quad (8.1)$$

contains, within the brackets, a term responsible for the incorporation of intraspecific competition. The basis of the Lotka–Volterra model is the replacement of this term by one which incorporates both intra- and interspecific competition.

The population size of one species can be denoted by N_1 , and that of a second species by N_2 . Their carrying capacities and intrinsic rates of increase are K_1 , K_2 , r_1 and r_2 , respectively.

Suppose that 10 individuals of species 2 have, between them, the same competitive, inhibitory effect on species 1 as does a single individual of

α : the competition coefficient

species 1. The total competitive effect on species 1 (intra- and interspecific) will then be equivalent to the effect of $(N_1 + N_2/10)$ species 1 individuals. The constant (1/10 in the present case) is called a competition coefficient and is denoted by α_{12} (‘alpha-one-two’). It measures the per capita competitive effect on species 1 of species 2. Thus, multiplying N_2 by α_{12} converts it to a number of ‘ N_1 -equivalents’. (Note that $\alpha_{12} < 1$ means that individuals of species 2 have less inhibitory effect on individuals of species 1 than individuals of species 1 have on others of their own species, whilst $\alpha_{12} > 1$ means that individuals of species 2 have a greater inhibitory effect on individuals of species 1 than do the species 1 individuals themselves.)

Lotka–Volterra model: a logistic model for two species

The crucial element in the model is the replacement of N_1 in the bracket of the logistic equation with a term signifying ‘ N_1 plus N_1 -equivalents’, i.e.:

$$\frac{dN_1}{dt} = r_1 N_1 \frac{(K_1 - (N_1 + \alpha_{12} N_2))}{K_1} \tag{8.2}$$

or:

$$\frac{dN_1}{dt} = r_1 N_1 \frac{(K_1 - N_1 - \alpha_{12} N_2)}{K_1} \tag{8.3}$$

and in the case of the second species:

$$\frac{dN_2}{dt} = r_2 N_2 \frac{(K_2 - N_2 - \alpha_{21} N_1)}{K_2} \tag{8.4}$$

These two equations constitute the Lotka–Volterra model.

behavior of the Lotka–Volterra model is investigated using ‘zero isoclines’

To appreciate the properties of this model, we must ask the question: when (under what circumstances) does each species increase or decrease in abundance? In order to answer this, it is necessary to construct diagrams in

which all possible combinations of species 1 and species 2 abundance can be displayed (i.e. all possible combinations of N_1 and N_2). These will be diagrams (Figures 8.7 and 8.9), with N_1 plotted on the horizontal axis and N_2 plotted on the vertical axis, such that there are low numbers of both species towards the bottom left, high numbers of both species towards the top right, and so on. Certain combinations of N_1 and N_2 will give rise to increases in species 1 and/or species 2, whilst other combinations will give rise to decreases in species 1 and/or species 2. Crucially, there

must also therefore be ‘zero isoclines’ for each species (lines along which there is neither an increase nor a decrease), dividing the combinations leading to increase from those leading to decrease. Moreover, if a zero isocline is drawn first, there will be combinations leading to an increase on one side of it, and combinations leading to a decrease on the other.

In order to draw a zero isocline for species 1, we can use the fact that on the zero isocline $dN_1/dt = 0$ (by definition), that is (from Equation 8.3):

$$r_1 N_1 (K_1 - N_1 - \alpha_{12} N_2) = 0. \tag{8.5}$$

This is true when the intrinsic rate of increase (r_1) is zero, and when the population size (N_1) is zero, but – much more importantly in the present context – it is also true when:

$$K_1 - N_1 - \alpha_{12} N_2 = 0, \tag{8.6}$$

which can be rearranged as:

$$N_1 = K_1 - \alpha_{12} N_2. \tag{8.7}$$

In other words, everywhere along the straight line which this equation represents, $dN_1/dt = 0$. The line is therefore the zero isocline for species 1; and since it is a straight line it can be drawn by finding two points on it and joining them. Thus, in Equation 8.7, when:

$$N_1 = 0, N_2 = \frac{K_1}{\alpha_{12}} \text{ (point A, Figure 8.7a)} \tag{8.8}$$

and when:

$$N_2 = 0, N_1 = K_1 \text{ (point B, Figure 8.7a),} \tag{8.9}$$

and joining them gives the zero isocline for species 1. Below and to the left of this, the numbers of both species are relatively low, and species 1, subjected to only weak competition, increases in abundance (the arrows in the figure, representing this increase, point from left to right, since N_1 is on the horizontal axis). Above and to the right of the line, the numbers are high, competition is strong and species 1 decreases in abundance (arrows from right to left). Based on an equivalent derivation, Figure 8.7b has combinations leading to an increase and decrease in species 2, separated by a species 2 zero isocline, with arrows, like the N_2 axis, running vertically.

Finally, in order to determine the outcome of competition in this model, it is necessary to fuse Figures 8.7a and b, allowing the behavior of a joint population to be predicted. In doing this, it should be noted that the arrows in Figure 8.7 are actually vectors – with a strength as well as a direction – and that to determine

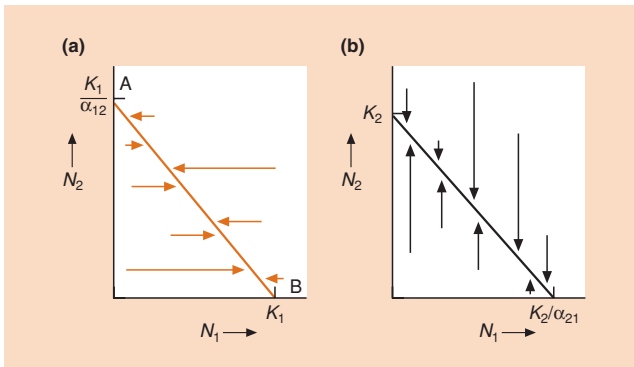


Figure 8.7 The zero isoclines generated by the Lotka–Volterra competition equations. (a) The N_1 zero isocline: species 1 increases below and to the left of it, and decreases above and to the right of it. (b) The equivalent N_2 zero isocline.

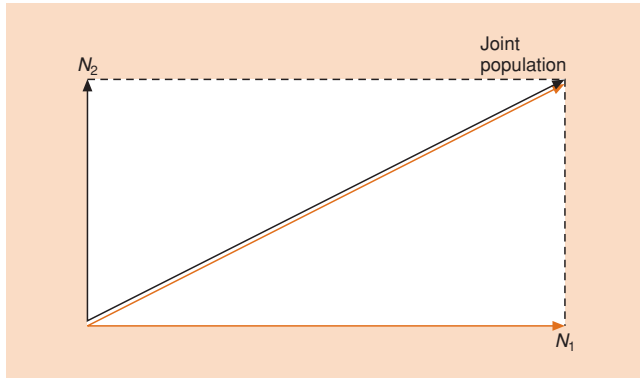


Figure 8.8 Vector addition. When species 1 and 2 increase in the manner indicated by the N_1 and N_2 arrows (vectors), the joint population increase is given by the vector along the diagonal of the rectangle, generated as shown by the N_1 and N_2 vectors.

four ways in which the two zero isoclines can be arranged

the behavior of a joint N_1, N_2 population, the normal rules of vector addition should be applied (Figure 8.8).

Figure 8.9 shows that there are, in fact, four different ways in which the

two zero isoclines can be arranged relative to one another, and the outcome of competition will be different in each case. The different cases can be defined and distinguished by the intercepts of the zero isoclines. For instance, in Figure 8.9a:

$$\frac{K_1}{\alpha_{12}} > K_2 \quad \text{and} \quad K_1 > \frac{K_2}{\alpha_{21}} \quad (8.10)$$

i.e.:

$$K_1 > K_2\alpha_{12} \quad \text{and} \quad K_1\alpha_{21} > K_2. \quad (8.11)$$

The first inequality ($K_1 > K_2\alpha_{12}$) indicates that the inhibitory intraspecific effects that species 1 can exert on itself are greater than the interspecific effects that species 2 can exert on species 1. The second inequality, however, indicates that species 1 can exert more of an effect on species 2 than species 2 can on itself. Species 1 is thus a strong interspecific competitor, whilst species 2 is a weak interspecific competitor; and as the vectors in Figure 8.9a show, species 1 drives species 2 to extinction and attains its own carrying capacity. The situation is

strong interspecific competitors outcompete weak interspecific competitors

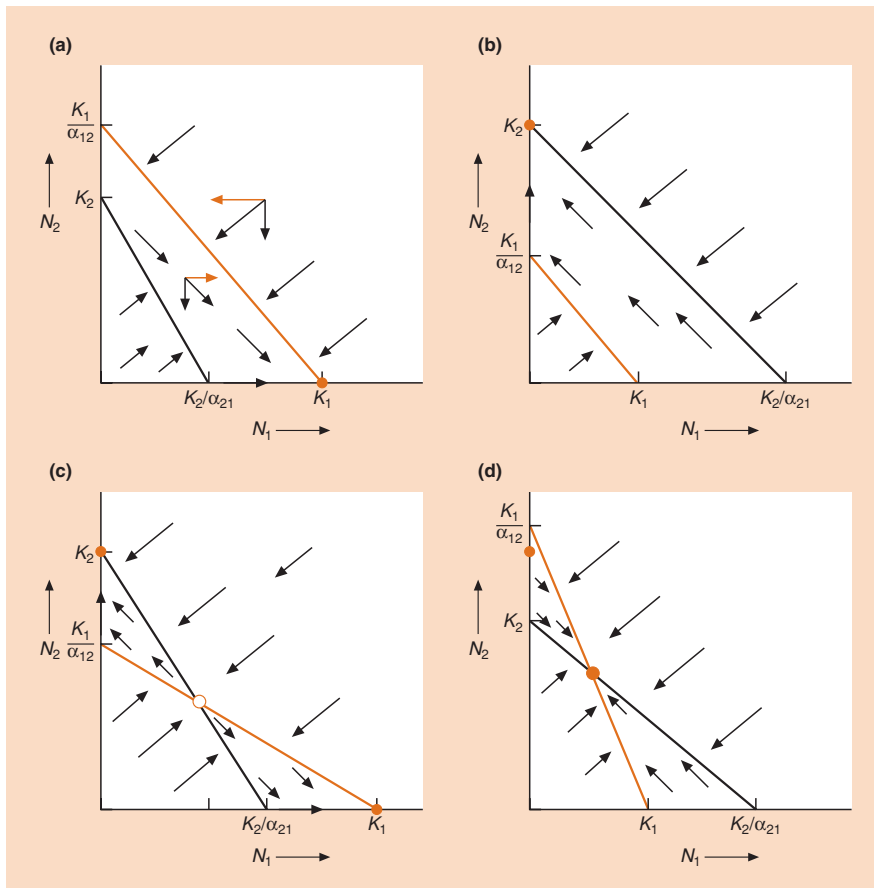


Figure 8.9 The outcomes of competition generated by the Lotka–Volterra competition equations for the four possible arrangements of the N_1 and N_2 zero isoclines. Vectors, generally, refer to joint populations, and are derived as indicated in (a). The solid circles show stable equilibrium points. The open circle in (c) is an unstable equilibrium point. For further discussion, see the text.

reversed in Figure 8.8b. Hence, Figures 8.8a and b describe cases in which the environment is such that one species invariably outcompetes the other.

In Figure 8.9c:

$$K_2 > \frac{K_1}{\alpha_{12}} \quad \text{and} \quad K_1 > \frac{K_2}{\alpha_{21}} \quad (8.12)$$

i.e.:

$$K_2\alpha_{12} > K_1 \quad \text{and} \quad K_1\alpha_{21} > K_2. \quad (8.13)$$

when interspecific competition is more important than intraspecific, the outcome depends on the species' densities

Thus, individuals of both species compete more strongly with individuals of the other species than they do amongst themselves. This will occur, for example, when each species produces a substance that is toxic to the other species but is harmless to itself, or

when each species is aggressive towards or even preys upon individuals of the other species, more than individuals of its own species. The consequence, as the figure shows, is an unstable equilibrium combination of N_1 and N_2 (where the isoclines cross), and two stable points. At the first of these stable points, species 1 reaches its carrying capacity with species 2 extinct; whilst at the second, species 2 reaches its carrying capacity with species 1 extinct. Which of these two outcomes is actually attained is determined by the initial densities: the species which has the initial advantage will drive the other species to extinction.

Finally, in Figure 8.9d:

$$\frac{K_1}{\alpha_{12}} > K_2 \quad \text{and} \quad \frac{K_2}{\alpha_{21}} > K_1 \quad (8.14)$$

i.e.:

$$K_1 > K_2\alpha_{12} \quad \text{and} \quad K_2 > K_1\alpha_{21}. \quad (8.15)$$

when interspecific competition is less important than intraspecific, the species coexist

In this case, both species have less competitive effect on the other species than they have on themselves. The outcome, as Figure 8.9d shows, is a stable equilibrium combination of the two species, which all joint populations tend to approach.

Overall, therefore, the Lotka–Volterra model of interspecific competition is able to generate a range of possible outcomes: the predictable exclusion of one species by another, exclusion dependent on initial densities, and stable coexistence. Each of these possibilities will be discussed in turn, alongside the results of laboratory and field investigations. We will see that the three outcomes from the model correspond to biologically reasonable

circumstances. The model, therefore, in spite of its simplicity and its failure to address many of the complexities of the dynamics of competition in the real world, serves a useful purpose.

Before we move on, however, one particular shortcoming of the Lotka–Volterra model is worth noting. The

Ks, α s and r s

outcome of competition in the model depends on the K s and the α s, but not on the r s, the intrinsic rates of increase. These determine the speed with which the outcome is achieved but not the outcome itself. This, though, seems to be a result peculiar to competition between only two species, since in models of competition between three or more species, the K s, α s and r s combine to determine the outcome (Strobeck, 1973).

8.4.2 The Competitive Exclusion Principle

Figure 8.9a and b describes cases in which a strong interspecific competitor invariably outcompetes a weak inter-

fundamental and realized niches

specific competitor. It is useful to consider this situation from the point of view of niche theory (see Sections 2.2 and 3.8). Recall that the niche of a species in the absence of competition from other species is its *fundamental* niche (defined by the combination of conditions and resources that allow the species to maintain a viable population). In the presence of competitors, however, the species may be restricted to a *realized* niche, the precise nature of which is determined by which competing species are present. This distinction stresses that interspecific competition reduces fecundity and survival, and that there may be parts of a species' fundamental niche in which, as a result of interspecific competition, the species can no longer survive and reproduce successfully. These parts of its fundamental niche are absent from its realized niche. Thus, returning to Figures 8.9a and b, we can say that the weak interspecific competitor lacks a realized niche when in competition with the stronger competitor. The real examples of interspecific competition previously discussed can now be re-examined in terms of niches.

In the case of the diatom species, the fundamental niches of both species were provided by the laboratory regime (they both thrived when alone). Yet when *Synedra* and *Asterionella* competed, *Synedra* had a realized niche whilst *Asterionella* did not: there was competitive exclusion of *Asterionella*.

coexisting competitors often exhibit a differentiation of their realized niches

The same outcome was recorded when Gause's *P. aurelia* and *P. caudatum* competed; *P. caudatum* lacked a realized niche and was competitively excluded by *P. aurelia*. When *P. caudatum* and *P. bursaria* competed, on the other hand, both species had realized niches, but these niches were noticeably different: *P. caudatum* living and feeding on the bacteria in the medium, *P. bursaria* concentrating on the yeast cells on the bottom of the

tube. Coexistence was therefore associated with a differentiation of realized niches, or a ‘partitioning’ of resources.

In the *Galium* experiments, the fundamental niches of both species included both acidic and calcareous soils. In competition with one another, however, the realized niche of *G. hercynicum* was restricted to acidic soils, whilst that of *G. pumilum* was restricted to calcareous ones – there was reciprocal competitive exclusion. Neither habitat allowed niche differentiation, and neither habitat fostered coexistence.

Amongst Taniguchi and Nakano’s salmonid fishes, the fundamental niches of each species extended over a broad range in altitude (and temperature) but both were restricted to a smaller realized niche (Dolly Varden at higher altitudes and white-spotted charr at lower altitudes).

Similarly, amongst Connell’s barnacles, the fundamental niche of *Chthamalus* extended down into the *Balanus* zone, but competition from *Balanus* restricted *Chthamalus* to a realized niche higher up the shore. In other words, *Balanus* competitively excluded *Chthamalus* from the lower zones, but for *Balanus* itself, even its fundamental niche did not extend up into the *Chthamalus* zone: its sensitivity to desiccation prevented it surviving even in the absence of *Chthamalus*. Hence, overall, the coexistence of these species was also associated with a differentiation of realized niches.

the Competitive Exclusion Principle

The pattern that has emerged from these examples has also been uncovered in many others, and has been elevated to the status of a principle: the *Competitive Exclusion Principle* or ‘Gause’s Principle’. It can be stated as follows: if two competing species coexist in a stable environment, then they do so as a result of niche differentiation, i.e. differentiation of their realized niches. If, however, there is no such differentiation, or if it is precluded by the habitat, then one competing species will eliminate or exclude the other. Thus exclusion occurs when the realized niche of the superior competitor completely fills those parts of the inferior competitor’s fundamental niche that are provided by the habitat.

difficulty proving and, especially, disproving the Principle

When there is coexistence of competitors, a differentiation of realized niches is sometimes seen to arise from current competition (an ‘ecological’ effect), as with the barnacles. Often, however, the niche differentiation is believed to have arisen either as a result of the past elimination of those species without realized niches (leaving behind only those exhibiting niche differentiation – another ecological effect) or as an *evolutionary* effect of competition. In either case, present competition may be negligible or at least impossible to detect. Consider again the coexisting tits. The species coexist and exhibit differentiation of their realized niches. But we do not know whether they compete now, or have ever competed in the past, or whether other species have been competitively excluded in the

past. It is impossible to say with certainty whether the Competitive Exclusion Principle was relevant. If the species do actually compete currently, or if other species are being or have been competitively excluded, then the Principle is relevant in the strictest sense. If they competed only in the past, and that competition has led to their niche differentiation, then the Principle is relevant, but only if it is extended from applying to the coexistence of ‘competitors’ to the coexistence of ‘species that are *or have ever been* competitors’. Of course, if the species have never competed, then the Principle is of no relevance here. Clearly, interspecific competition cannot be studied by the mere documentation of present interspecific differences.

With Martin and Martin’s warblers, on the other hand, the two species competed and coexisted, and the Competitive Exclusion Principle would suggest that this was a result of niche differentiation. But, whilst reasonable, this is by no means proven, since such differentiation was neither observed nor shown to be effective. Thus, when two competitors coexist, it is often difficult to establish positively that there is niche differentiation. Worse still, it is impossible to prove the absence of it. When ecologists fail to find differentiation, this might simply mean that they have looked in the wrong place or in the wrong way. Clearly, there can be very real methodological problems in establishing the pertinence of the Competitive Exclusion Principle in any particular case.

The Competitive Exclusion Principle has become widely accepted because: (i) there is much good evidence in its favor; (ii) it makes intuitive good sense; and (iii) there are theoretical grounds for believing in it (the Lotka–Volterra model). But there will always be cases in which it has not been positively established; and as Section 8.5 will make plain, there are many other cases in which it simply does not apply. In short, interspecific competition is a process that is often associated, ecologically and evolutionarily, with a particular pattern (niche differentiation), but interspecific competition and niche differentiation (the process and the pattern) are not inextricably linked. Niche differentiation can arise through other processes, and interspecific competition need not lead to a differentiation of niches.

niche differentiation and interspecific competition: a pattern and a process not always linked

8.4.3 Mutual antagonism

Figure 8.9c, derived from the Lotka–Volterra model, describes a situation in which interspecific competition is, for both species, a more powerful force than intraspecific competition. This is known as mutual antagonism.

An extreme example of such a situation is provided by work on two species of flour beetle: *Tribolium confusum* and *T. castaneum* (Park, 1962). Park’s experiments in the 1940s,

reciprocal predation in flour beetles

Table 8.1 Reciprocal predation (a form of mutual antagonism) between two species of flour beetle, *Tribolium confusum* and *T. castaneum*. Both adults and larvae eat both eggs and pupae. In each case, and overall, the preference of each species for its own or the other species is indicated. Interspecific predation is more marked than intraspecific predation. (After Park *et al.*, 1965.)

	'Predator'	'Shows a preference for ...'
Adults eating eggs	<i>T. confusum</i>	<i>T. confusum</i>
	<i>T. castaneum</i>	<i>T. confusum</i>
Adults eating pupae	<i>T. confusum</i>	<i>T. castaneum</i>
	<i>T. castaneum</i>	<i>T. confusum</i>
Larvae eating eggs	<i>T. confusum</i>	<i>T. castaneum</i>
	<i>T. castaneum</i>	<i>T. castaneum</i>
Larvae eating pupae	<i>T. confusum</i>	<i>T. castaneum</i>
	<i>T. castaneum</i>	<i>T. confusum</i>
Overall	<i>T. confusum</i>	<i>T. castaneum</i>
	<i>T. castaneum</i>	<i>T. confusum</i>

1950s and 1960s were amongst the most influential in shaping ideas about interspecific competition. He reared the beetles in simple containers of flour, which provided fundamental and often realized niches for the eggs, larvae, pupae and adults of both species. There was certainly exploitation of common resources by the two species; but in addition, the beetles preyed upon each other. The larvae and adults ate eggs and pupae, cannibalizing their own species as well as attacking the other species, and their propensity for doing so is summarized in Table 8.1. The important point is that taken overall, beetles of both species ate more individuals of the other species than they did of their own. Thus, a crucial mechanism in the interaction of these competing species was reciprocal predation (i.e. mutual antagonism), and it is easy to see that both species were more affected by inter- than intraspecific predation.

the outcome is probable rather than definite

Figure 8.9c, the Lotka–Volterra model, suggests that the consequences of mutual antagonism are essentially the same whatever the exact mechanism. Because species are affected more

by inter- than intraspecific competition, the outcome is strongly dependent on the relative abundances of the competing species. The small amount of interspecific aggression displayed by a rare species will have relatively little effect on an abundant competitor; but the large amount of aggression displayed by an abundant species might easily drive a rare species to local extinction. Moreover, if abundances are finely balanced, a small change in relative abundance will be sufficient to shift the advantage from one species to the other. The outcome of competition will then be unpredictable – either species could exclude the other, depending on the exact densities that they start with or attain.

Table 8.2 Competition between *Tribolium confusum* and *T. castaneum* in a range of climates. One species is always eliminated and climate alters the outcome, but at intermediate climates the outcome is nevertheless probable rather than definite. (After Park, 1954.)

Climate	Percentage wins	
	<i>T. confusum</i>	<i>T. castaneum</i>
Hot–moist	0	100
Temperate–moist	14	86
Cold–moist	71	29
Hot–dry	90	10
Temperate–dry	87	13
Cold–dry	100	0

Table 8.2 shows that this was indeed the case with Park’s flour beetles. There was always only one winner, and the balance between the species changed with climatic conditions. Yet at all intermediate climates *the outcome was probable rather than definite*. Even the inherently inferior competitor occasionally achieved a density at which it could outcompete the other species.

8.5 Heterogeneity, colonization and preemptive competition

At this point it is necessary to sound a loud note of caution. It has been assumed in this chapter until now that the environment is sufficiently constant for the outcome of competition to be determined by the competitive abilities of the competing species. In reality, though, such situations are far from universal. Environments are usually a patchwork of favorable and unfavorable habitats; patches are often only available temporarily; and patches often appear at unpredictable times and in unpredictable places. Even when interspecific competition occurs, it does not necessarily continue to completion. Systems do not necessarily reach equilibrium, and superior competitors do not necessarily have time to exclude their inferiors. Thus, an understanding of interspecific competition itself is not always enough. It is often also necessary to consider how interspecific competition is influenced by, and interacts with, an inconstant or unpredictable environment. To put it another way: K_s and α_s alone may determine an equilibrium, but in nature, equilibria are very often not achieved. Thus, the speed with which an equilibrium is approached becomes important. That is, as we have already noted in Section 8.4.1 in another context, not only K_s and α_s , but r_s too play their part.

a note of caution: competition is influenced by heterogeneous, inconstant or unpredictable environments

8.5.1 Unpredictable gaps: the poorer competitor is a better colonizer

'Gaps' of unoccupied space occur unpredictably in many environments. Fires, landslips and lightning can create gaps in woodlands; storm-force seas can create gaps on the shore; and voracious predators can create gaps almost anywhere. Invariably, these gaps are recolonized. But the first species to do so is not necessarily the one that is best able to exclude other species in the long term. Thus, so long as gaps are created at the appropriate frequency, it is possible for a 'fugitive' species and a highly competitive species to coexist. The fugitive species tends to be the first to colonize gaps; it establishes itself, and it reproduces. The other species tends to be slower to invade the gaps, but having begun to do so, it outcompetes and eventually excludes the fugitive from that particular gap.

fugitive annuals
and competitive
perennials

This outline sketch has been given some quantitative substance in a simulation model in which the 'fugitive' species is thought of as an annual plant and the superior competitor as a perennial (Crawley & May, 1987). The model is one of a growing number that combine temporal and spatial dynamics by having interactions occur within individual cells of a two-dimensional lattice, but also having movement between cells (see also Inghe, 1989; Dytham, 1994; Bolker *et al.*, 2003). In this model, each cell can either be empty or occupied by either a single individual of the annual or a single ramet of the perennial. Each 'generation', the perennial can invade cells adjacent to those it already occupies, and it does so irrespective of whether those cells support an annual (a reflection of the perennial's competitive superiority), but individual ramets of the perennial may also die. The annual, however, can colonize any empty cell, which it does through the deposition of randomly dispersed 'seed', the quantity of which reflects the annual's abundance. Putting details aside, the annual can coexist with its superior competitor, providing the product (cE^*) of the annual's fecundity (c) and the equilibrium proportion of empty cells (E^*) is sufficiently great (Figure 8.10), i.e. as long as the annual is a sufficiently good colonizer and there are sufficient opportunities for it to do so. Indeed, the greater cE^* , the more the balance in the equilibrium mixture shifts towards the annual (Figure 8.10).

An example is provided by the coexistence of the sea palm *Postelsia palmaeformis* (a brown alga) and the mussel *Mytilus californianus* on the coast of Washington (Paine, 1979). *Postelsia* is an annual that must re-establish itself each year in order to persist at a site. It does so by attaching to the bare rock, usually in gaps in the mussel bed created by wave action. However, the mussels themselves slowly encroach on these gaps, gradually filling them and precluding colonization by

coexistence of a
competitive mussel
and a fugitive sea
palm

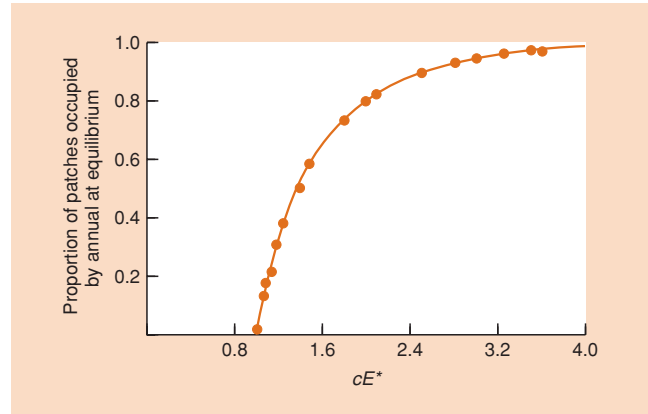


Figure 8.10 In a spatial lattice, a model fugitive annual plant can coexist with a competitively superior perennial provided $cE^* > 1$ (where c is the annual's fecundity and E^* the equilibrium proportion of empty cells in the lattice). For larger values, the fraction of cells occupied by the annual increases with cE^* . (After Crawley & May, 1987.)

Postelsia. Paine found that these species coexisted only at sites in which there was a relatively high average rate of gap formation (about 7% of surface area per year), and in which this rate was approximately the same each year. Where the average rate was lower, or where it varied considerably from year to year, there was (either regularly or occasionally) a lack of bare rock for colonization. This led to the overall exclusion of *Postelsia*. At the sites of coexistence, on the other hand, although *Postelsia* was eventually excluded from each gap, these were created with sufficient frequency and regularity for there to be coexistence in the site as a whole.

8.5.2 Unpredictable gaps: the preemption of space

When two species compete on equal terms, the result is usually predictable. But in the colonization of unoccupied

first come, first
served

space, competition is rarely even handed. Individuals of one species are likely to arrive, or germinate from the seed bank, in advance of individuals of another species. This, in itself, may be enough to tip the competitive balance in favor of the first species. If space is preempted by different species in different gaps, then this may allow coexistence, even though one species would always exclude the other if they competed 'on equal terms'.

For instance, Figure 8.11 shows the results of a competition experiment between the annual grasses *Bromus madritensis* and *B. rigidus*, which occur together in Californian rangelands (Harper, 1961). When they were sown simultaneously in an equiproportional mixture, *B. rigidus* contributed overwhelmingly to the biomass of the mixed population. But, by delaying the introduction of *B. rigidus*

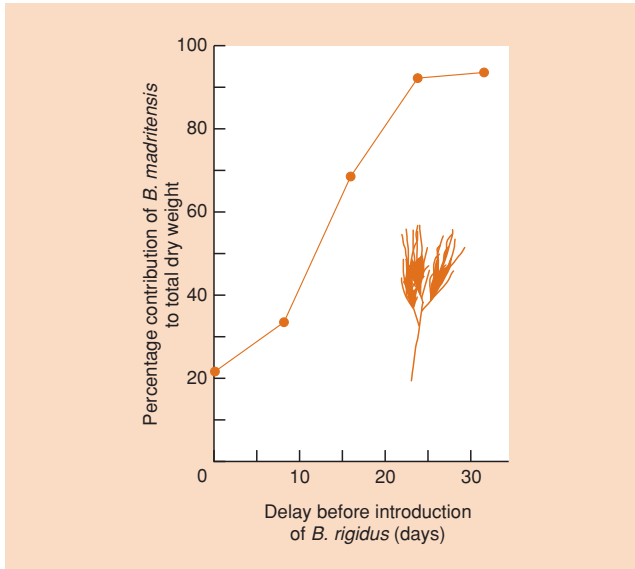


Figure 8.11 The effect of timing on competition. *Bromus rigidus* makes an overwhelming contribution to the total dry weight per pot after 126 days growth when sown at the same time as *B. madritensis*. But, as the introduction of *B. rigidus* is delayed, its contribution declines. Total yield per pot was unaffected by delaying the introduction of *B. rigidus*. (After Harper, 1961.)

into the mixtures, the balance was tipped decisively in favour of *B. madritensis*. It is therefore quite wrong to think of the outcome of competition as being always determined by the inherent competitive abilities of the competing species. Even an ‘inferior’ competitor can exclude its superior if it has enough of a head start. This can foster coexistence when repeated colonization occurs in a changing or unpredictable environment.

8.5.3 Fluctuating environments

paradox of the plankton

The balance between competing species can be shifted repeatedly, in fact, and coexistence therefore fostered, simply as a result of environmental change. This was the argument used by Hutchinson (1961) to explain the ‘paradox of the plankton’ – the paradox being that numerous species of planktonic algae frequently coexist in simple environments with little apparent scope for niche differentiation. Hutchinson suggested that the environment, although simple, was continually changing, particularly on a seasonal basis. Thus, although the environment at any one time would tend to promote the exclusion of certain species, it would alter and perhaps even favor these same species before exclusion occurred. In other words, the equilibrium outcome of a competitive interaction may not be of paramount importance

if the environment typically changes long before the equilibrium can be reached.

8.5.4 Ephemeral patches with unpredictable lifespans

Many environments, by their very nature, are not simply variable but ephemeral. Amongst the more obvious examples are decaying corpses (carrion), dung, rotting fruit and fungi, and temporary ponds. But note too that a leaf or an annual plant can be seen as an ephemeral patch, especially if it is palatable to its consumer for only a limited period. Often, these ephemeral patches have an unpredictable lifespan – a piece of fruit and its attendant insects, for instance, may be eaten at any time by a bird. In these cases, it is easy to imagine the coexistence of two species: a superior competitor and an inferior competitor that reproduces early.

coexistence of the strong with the fast . . .

One example concerns two species of pulmonate snail living in ponds in northeastern Indiana. Artificially altering the density of one or other species in the field showed that the fecundity of *Physa gyrina* was significantly reduced by interspecific competition from *Lymnaea elodes*, but the effect was not reciprocated. *L. elodes* was clearly the superior competitor when competition continued throughout the summer. Yet *P. gyrina* reproduced earlier and at a smaller size than *L. elodes*, and in the many ponds that dried up by early July it was often the only species to have produced resistant eggs in time. The species therefore coexisted in the area as a whole, in spite of *P. gyrina*’s apparent inferiority (Brown, 1982). Among frogs and toads, on the other hand, the competitively superior tadpoles of *Scaphiopus holbrookii* are even more successful when ponds dry up because they have shorter larval periods than weaker competitors such as *Hyla chrysoscelis* (Wilbur, 1987).

. . . but not always

8.5.5 Aggregated distributions

A more subtle, but more generally applicable path to the coexistence of a superior and an inferior competitor on a patchy and ephemeral resource is based on the idea that the two species may have independent, aggregated (i.e. clumped) distributions over the available patches. This would mean that the powers of the superior competitor were mostly directed against members of its own species (in the high-density clumps), but that this aggregated superior competitor would be absent from many patches – within which the inferior competitor could escape competition. An inferior competitor may then be able to coexist with a superior competitor that would rapidly exclude it

a clumped superior competitor adversely affects itself and leaves gaps for its inferior

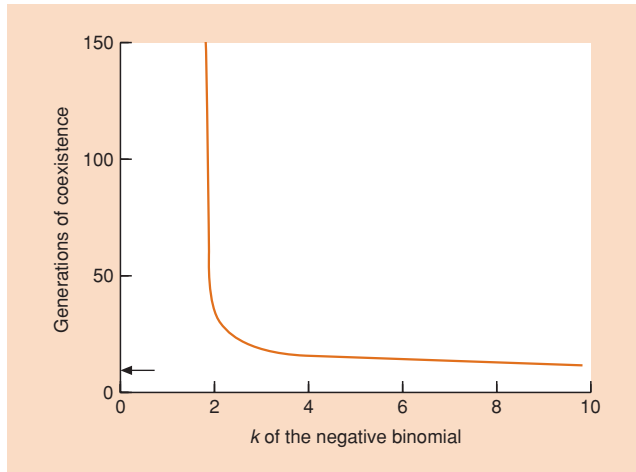


Figure 8.12 When two species compete on a continuously distributed resource, one species would exclude the other in approximately 10 generations (as indicated by the arrow). However, with these same species on a patchy and ephemeral resource, the number of generations of coexistence increases with the degree of aggregation of the competitors, as measured by the parameter k of the ‘negative binomial’ distribution. Values above 5 are effectively random distributions; values below 5 represent increasingly aggregated distributions. (After Atkinson & Shorrocks, 1981.)

from a continuous, homogeneous environment. Certainly it can do so in models (see, for example, Atkinson & Shorrocks, 1981; Kreitman *et al.*, 1992; Dieckmann *et al.*, 2000). For instance, a simulation model (Figure 8.12) shows that the persistence of such coexistence between competitors increases with the degree of aggregation (as measured by the parameter k of the ‘negative binomial’ distribution) until, at high levels of aggregation, coexistence is apparently permanent, although this has nothing to do with any niche differentiation. Since many species have aggregated distributions in nature, these results may be applicable widely.

Note, however, that whilst such coexistence of competitors has nothing to do with niche differentiation, it is linked to it by a common theme – that of species competing more frequently and intensively intraspecifically than they do interspecifically. Niche differentiation is one means by which this can occur, but temporary aggregations can give rise to the same phenomenon – even for the inferior competitor.

In seeking to justify the applicability of these models to the real world, however, one question in particular needs to be answered: are two similar species really likely to have independent distributions over available patches of resource? The question has been addressed through an examination of a large number of data sets from Diptera, especially drosophilid flies – where eggs are laid, and larvae develop, in ephemeral patches (fruits, fungi, flowers, etc.). In fact, there was little evidence for independence

in the aggregations of coexisting species (Shorrocks *et al.*, 1990; see also Worthen & McGuire, 1988). However, computer simulations suggest that whilst a positive association between species (i.e. a tendency to aggregate in the same patches) does make coexistence more difficult, the level of association and aggregation actually found would still generally lead to coexistence, whereas there would be exclusion in a homogeneous environment (Shorrocks & Rosewell, 1987).

The importance of aggregation for coexistence has been further supported by another spatially explicit model based on a two-dimensional lattice of cells (see Section 8.5.1), each of which could be occupied by one of five species of grass: *Agrostis stolonifera*, *Cynosurus cristatus*, *Holcus lanatus*, *Lolium perenne* and *Poa trivialis* (Silvertown *et al.*, 1992). The model was a ‘cellular automaton’, in which each cell can exist in a limited number of discrete states (in this case, which species was in occupancy), with the state of each cell determined at each time step by a set of rules. In this case, the rules were based on the cell’s current state, the state of the neighboring cells and the probability that a species in a neighboring cell would replace its current occupant. These replacement rates of each species by each other species were themselves based on field observations (Thórhallsdóttir, 1990).

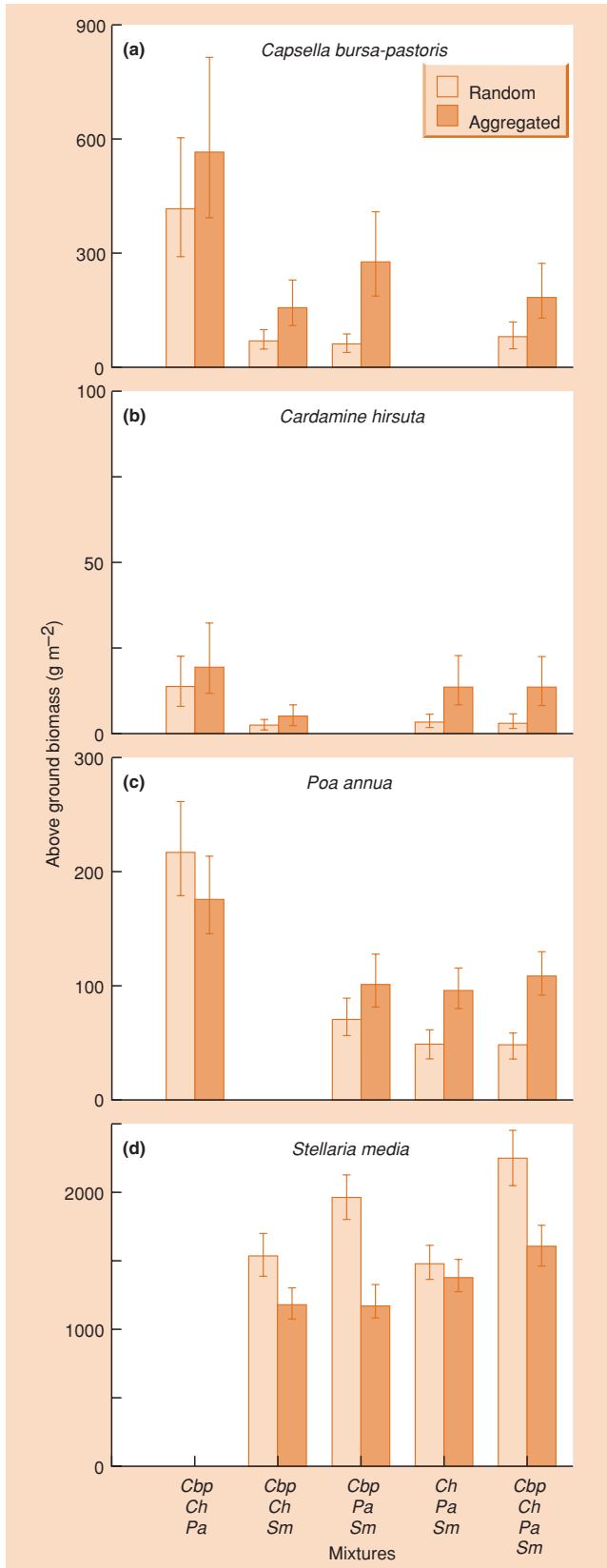
If the initial arrangement of the species over the grid was random (no aggregation), the three competitively inferior species were quickly driven to extinction, and of the survivors, *Agrostis* (greater than 80% cell occupancy) rapidly dominated *Holcus*. If, however, the initial arrangement was five equally broad single-species bands across the landscape, the outcome changed dramatically: (i) competitive exclusion was markedly delayed even for the worst competitors (*Cynosurus* and *Lolium*); (ii) *Holcus* sometimes occupied more than 60% of the cells, at a time (600 time steps) where, with an initially random arrangement, it would have been close to extinction; and (iii) the outcome itself depended largely on which species started next to each other, and hence, initially competed with each other.

There is no suggestion, of course, that natural communities of grasses exist as broad single-species bands – but, neither are we likely to find communities with species mixed at random, such that there is no spatial organization to be taken into account. The model emphasizes the dangers of ignoring aggregations (because they shift the balance towards intra- rather than interspecific competition, and hence promote coexistence), but also the dangers of ignoring the juxtaposition of aggregations, since these too may serve to keep competitive subordinates away from their superiors.

Despite a rich body of theory and models, there are few experimental studies that directly address the impact of spatial patterns on population dynamics. Stoll and Prati (2001) performed experiments with real plants in a study that had much in common with Silvertown’s

grasses in a cellular automaton

plants in a field experiment



theoretical treatment. They tested the hypothesis that intraspecific aggregation can promote coexistence and thus maintain high species richness in experimental communities of four annual terrestrial plants: *Capsella bursa-pastoris*, *Cardamine hirsuta*, *Poa annua* and *Stellaria media*. *Stellaria* is known to be the superior competitor among these species. Replicate three- and four-species mixtures were sown at high density, and the seeds were either placed completely at random or seeds of each species were aggregated in subplots within the experimental areas. Intraspecific aggregation decreased the performance of the superior *Stellaria* in the mixtures, whereas in all but one case aggregation improved the performance of the three inferior competitors (Figure 8.13).

More generally, the success of ‘neighborhood’ approaches (Pacala, 1997) in the study of plant competition, where the focus is on the competition experienced by individuals in local patches, rather than densities averaged out over whole populations, argues again in favor of the importance of acknowledging spatial heterogeneity. Coomes *et al.* (2002), for example, investigated competition between two species of sand-dune plant, *Aira praecox* and *Erodium cicutarium*, in northwest England. The smaller plant, *Aira*, tended to be aggregated even at the smallest spatial scales, whereas *Erodium* was moderately aggregated in patches of 30 and 50 mm radius but, if anything, was evenly spaced within 10 mm radius patches (Figure 8.14a). The two species, though, were negatively associated with one another at the smallest spatial scale (Figure 8.14b), indicating that *Aira* tended to occur in small, single-species clumps. *Aira* was therefore much less liable to competition from *Erodium* than would be the case if they were distributed at random, justifying the application by Coomes *et al.* of simulation models of competition where local responses were explicitly incorporated.

heterogeneity often stabilizes

Repeatedly in this section, then, the heterogeneous nature of the environment can be seen to have fostered coexistence without there being a marked differentiation of niches. A realistic view of interspecific competition, therefore, must acknowledge that it often proceeds not in isolation, but under the influence of, and within the constraints of, a patchy, impermanent or unpredictable world. Furthermore, the heterogeneity need not be in the temporal or spatial dimensions that we have discussed so far. Individual variation in competitive ability

Figure 8.13 (left) The effect of intraspecific aggregation on above-ground biomass (mean ± SE) of four plant species grown for 6 weeks in three- and four-species mixtures (four replicates of each). The normally competitively superior *Stellaria media* (Sm) did consistently less well when seeds were aggregated than when they were placed at random. In contrast, the three competitively inferior species – *Capsella bursa-pastoris* (Cbp), *Cardamine hirsuta* (Ch) and *Poa annua* (Pa) – almost always performed better when the seeds had been aggregated. Note the different scales on the vertical axes. (From Stoll & Prati, 2001.)

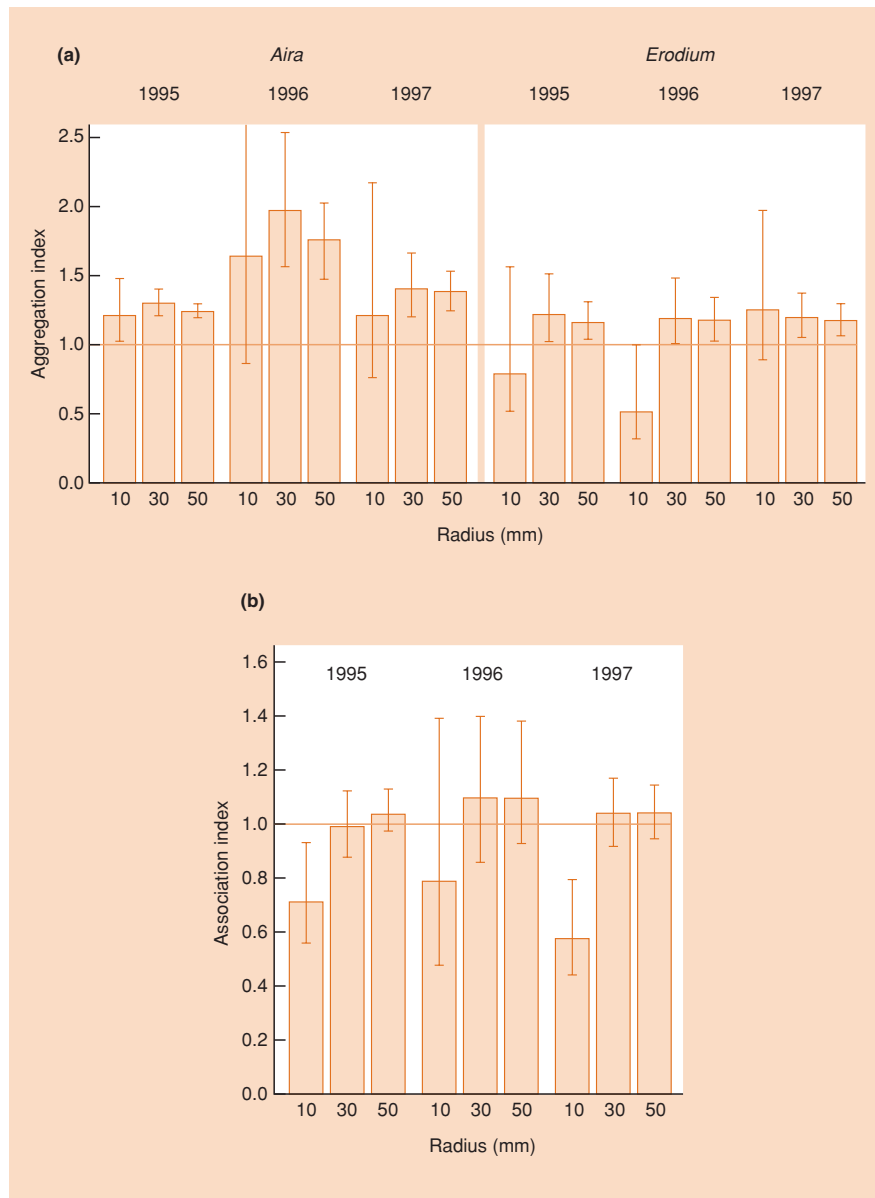


Figure 8.14 (a) Spatial distribution of two sand-dune species, *Aira praecox* and *Erodium cicutarium* at a site in northwest England. An aggregation index of 1 indicates a random distribution. Indices greater than 1 indicate aggregation (clumping) within patches with the radius as specified; values less than 1 indicate a regular distribution. Bars represent 95% confidence intervals. (b) The association between *Aira* and *Erodium* in each of the 3 years. An association index greater than 1 indicates that the two species tended to be found together more than would be expected by chance alone in patches with the radius as specified; values less than 1 indicate a tendency to find one species or the other. Bars represent 95% confidence intervals. (After Coomes *et al.*, 2002.)

within species can also foster stable coexistence in cases where a superior nonvariable competitor would otherwise exclude an inferior nonvariable species (Begon & Wall, 1987). This reinforces a point that recurs throughout this text: heterogeneity (spatial, temporal or individual) can have a stabilizing influence on ecological interactions.

8.6 Apparent competition: enemy-free space

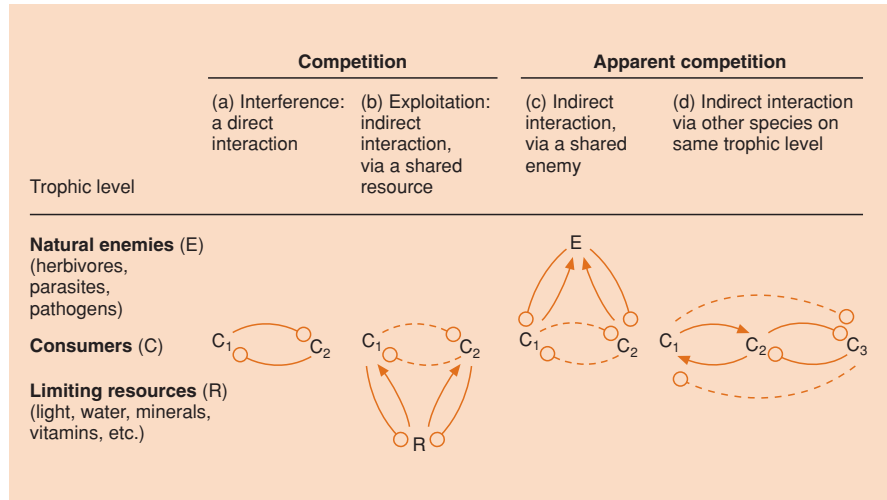
Another reason for being cautious in our discussion of competition is the existence of what Holt (1977, 1984) has called

‘apparent competition’, and what others have called ‘competition for enemy-free space’ (Jeffries & Lawton, 1984, 1985).

Imagine a single species of predator or parasite that attacks two species of prey (or host). Both prey species are harmed by the enemy, and the enemy benefits from both species of prey. Hence, the increase in abundance that the enemy achieves by consuming prey 1 increases the harm it does to prey 2. Indirectly, therefore, prey 1 adversely affects prey 2 and vice versa. These

two prey species attacked by a predator are, in essence, indistinguishable from two consumer species competing for a resource

Figure 8.15 In terms of the signs of their interactions, all of the following are indistinguishable from one another: (a) two species interfering directly (interference competition); (b) two species consuming a common resource (exploitation competition); (c) two species being attacked by a common predator ('apparent competition' for 'enemy-free space'); and (d) two species linked by a third which is a competitor of one and a mutualist of the other. (—), direct interactions; (---), indirect interactions; arrows indicate positive influences, circles indicate negative influences. (After Holt, 1984; Connell, 1990.)



interactions are summarized in Figure 8.15, which shows that from the point of view of the two prey species, the signs of the interactions are indistinguishable from those that would apply in the indirect interaction of two species competing for a single resource (exploitation competition). In the present case there appears to be no limiting resource. Hence, the term 'apparent competition'.

evidence for apparent competition . . .
 . . . in two caterpillars sharing a parasitoid,
 . . .

In an experiment involving a parasitoid (the ichneumonid wasp *Venturia canescens*) and two caterpillar hosts (*Plodia interpunctella* and *Ephestia kuehniella*), Bonsall and Hassell (1997) allowed free passage of the parasitoid between the host species but kept the

hosts apart to avoid the possibility of resource competition between them. When the experimental chambers contained just a single host species together with the parasitoid, both the parasite and host persisted and exhibited damped oscillations in population size, tending towards a stable equilibrium (Figure 8.16). But when the system was run with both host species, the parasitoid had a greater impact on the species with the lower intrinsic rate of increase (*E. kuehniella*). This host showed increasing population oscillations and invariably went extinct. By means of their elegant experimental design, Bonsall and Hassell were able to demonstrate the effect of apparent competition in a situation where resource competition between the caterpillar species was ruled out.

While the term 'apparent competition' is entirely appropriate, it is sometimes useful to think of 'enemy-free space' as the limiting resource for which prey (or host) species compete. This is because the persistence of prey species 1 will be favored by avoiding attacks from the predator, which we know also attacks prey 2. Clearly, prey 1 can achieve this by occupying a habitat, or adopting a form or a behavioral pattern, that is sufficiently different from that of prey 2. In short, 'being different' (i.e. niche

differentiation) will once again favor coexistence – but it will do so because it diminishes apparent competition or competition for enemy-free space.

A rare experimental demonstration of apparent competition for enemy-free space involves two groups of prey living on subtidal rocky reefs at Santa Catalina Island, California. The first comprises three species of mobile gastropods, *Tegula aureotincta*, *T. eiseni* and *Astraea undosa*; the second comprises sessile bivalves, dominated by the clam *Chama arcana*. Both groups were preyed upon by a lobster (*Panulirus interruptus*), an octopus (*Octopus bimaculatus*) and a whelk (*Kelletia kelletii*), although these predators showed a marked preference for the bivalves. In areas characterized by large boulders and much crevice space ('high relief') there were high densities of bivalves and predators, but only moderate densities of gastropods; whereas in low relief areas largely lacking crevice space ('cobble fields') there were apparently no bivalves, only a few predators but high densities of gastropods.

. . . in gastropods, bivalves and their predators . . .

The densities of the two prey groups were inversely correlated, but there was little in their feeding biology to suggest that they were competing for a shared food resource. On the other hand, when bivalves were experimentally introduced into cobble-field areas, the number of predators congregating there increased, the mortality rates of the gastropods increased (often observably associated with lobster or octopus predation) and the densities of the gastropods declined (Figures 8.17a, b). Experimental manipulation of the (mobile) gastropods proved impossible, but cobble sites with high densities of gastropods supported higher densities of predators, and had higher mortality rates of experimentally added bivalves than did sites with relatively low densities of gastropods (Figure 8.17c). On the rare high relief sites without *Chama* bivalves, predator densities were lower, and gastropod densities higher, than was normally the case (Figure 8.17d). It seems clear that each prey group adversely affected the other through an

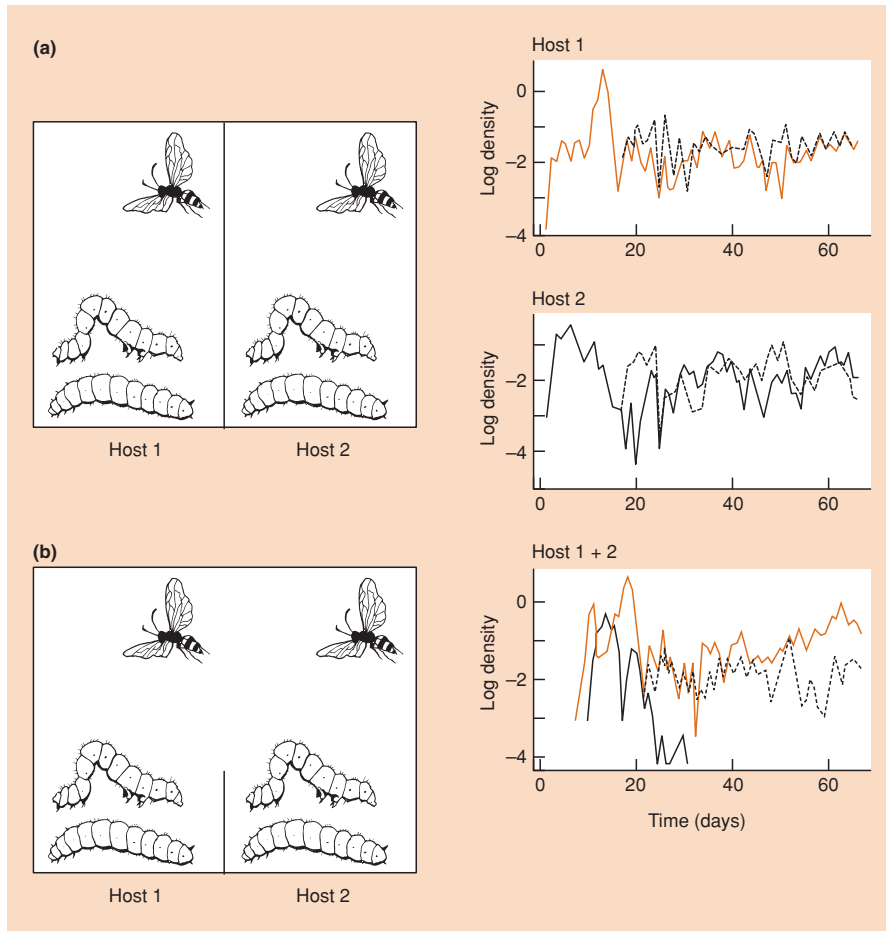


Figure 8.16 Parasite-mediated apparent competition via a parasitoid wasp *Venturia canescens* that lays eggs in two caterpillar host species. The experimental setups are illustrated on the left and the population dynamics of the parasitoid (dashed black lines) and host species (host 1 *Plodia interpunctella* (orange lines); host 2 *Ephestia kuehniella* (black lines)) on the right. (a) When only a single host was present, the parasitoid and host coexisted with stable dynamics. (b) When the parasitoid had access to both hosts, host 2 showed diverging oscillations and went extinct. (From Hudson & Greenman, 1998, after Bonsall & Hassell, 1997.)

increased number of predators, and hence increased predator-induced mortality.

... and in leaf-mining flies sharing parasitoids in a tropical forest

An experiment with a similar aim involved removing a common leaf-mining fly (*Calycomyza* sp.) and its host plant *Lepidaploa tortuosa* (Asteracea) in replicate sites in a tropical forest community in Belize, Central America.

Other leaf-mining fly species that shared natural enemies (parasitoid wasps) with *Calycomyza*, but whose host plants were different, demonstrated reduced parasitism and increased abundance (a year later) in the removal sites than in the control sites (Morris *et al.*, 2004). These results support predictions of apparent competition, involving a shared natural enemy, in a situation where interspecific competition among the fly species for host plants could not occur.

To complete the picture, there is another indirect interaction between two species that qualifies for the term 'apparent competition' (Figure 8.15d), where species 1 and 2 have negative impacts on one another, and species 2 and 3 have positive (mutualistic)

impacts (see Chapter 13). Species 1 and 3 then have indirect negative impacts on one another without sharing a common resource or, for that matter, a common predator. They exhibit apparent competition, although not for enemy-free space (Connell, 1990).

The examples mentioned so far concern apparent competition in animals. Connell (1990) carried out a particularly revealing reappraisal of 54 published plant examples of field experiments on 'competition', where the original authors had claimed to have demonstrated conventional interspecific competition in 50. A closer look revealed that, in many of these, insufficient information had been collected to distinguish between conventional competition and apparent competition; and in a number of others the information was available – but was ambiguous. For example, one study showed that removal of *Artemisia* bushes from a large site in Arizona led to much better growth of 22 species of herb than was observed in either undisturbed sites or sites where *Artemisia* was removed from narrow 3 m strips. This was originally interpreted in terms of greatly

reappraisal of plant competition

reappraisal of plant competition

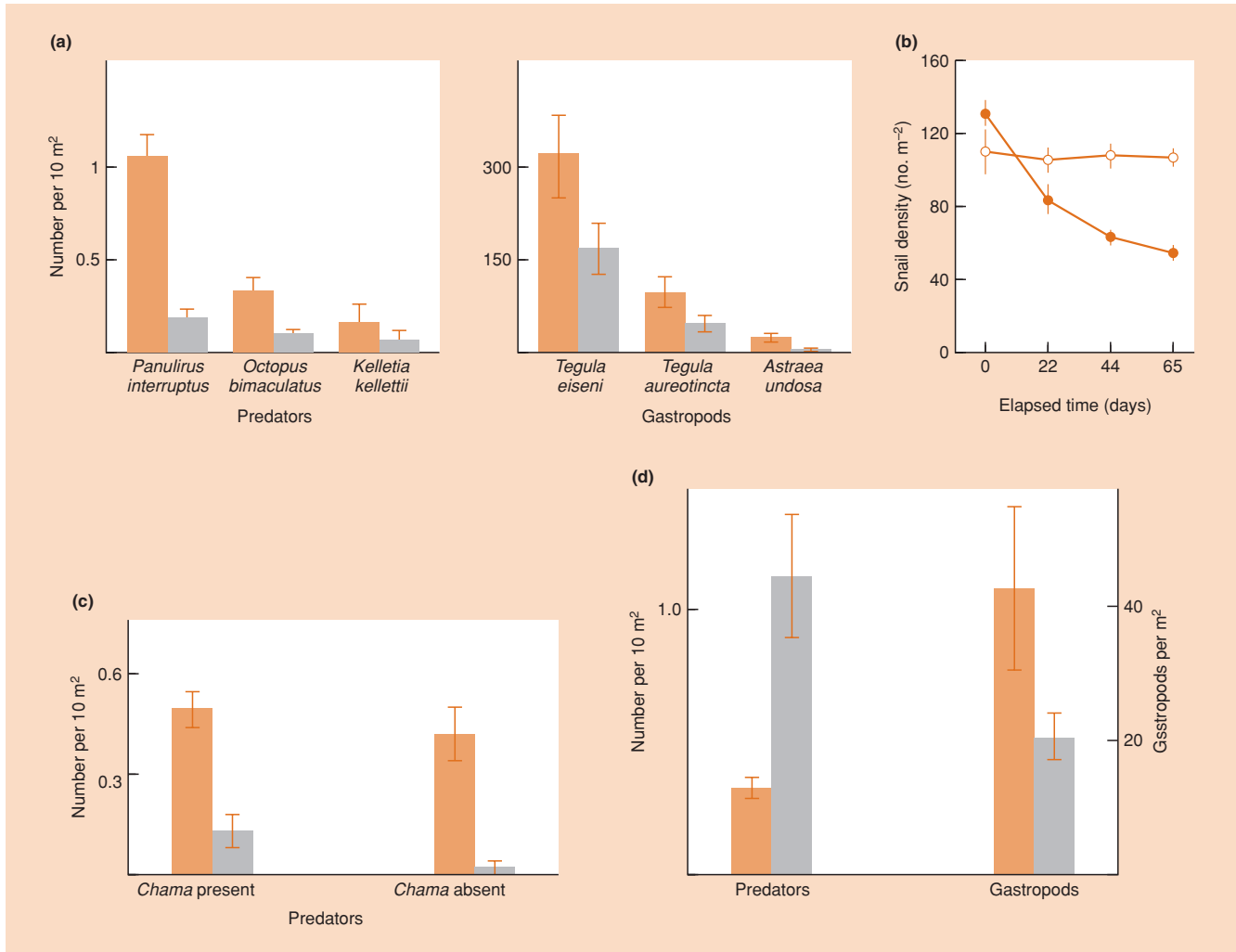


Figure 8.17 Evidence for apparent competition for predator-free space at Santa Catalina Island, USA. (a) Predator density (number per 10 m², with standard errors) and gastropod mortality increased (number of ‘newly dead’ shells per site, with standard errors) when bivalves were added to gastropod-dominated cobble sites (colored bars) relative to controls (gray bars). (b) This led to a decline in gastropod density (standard error bars shown). (c) Predator density was higher (number per 10 m², with standard errors) at high (colored bars) than at low (gray bars) gastropod-density cobble sites, both in the presence and absence of *Chama*. (d) Densities of predators were lower (number per 10 m², with standard errors) and densities of gastropods higher (number per m², with standard errors) at high-relief sites without *Chama* (colored bars) than at those with (gray bars). (After Schmitt, 1987.)

reduced exploitative competition for water in the former case (Robertson, 1947). However, the herbs in the larger site also experienced greatly reduced grazing pressure from deer, rodents and insects, for which the *Artemisia* bushes were not only a source of food but a place of shelter, too. The outcome is therefore equally likely to have resulted from reduced apparent competition.

distinguishing pattern and process

This emphasizes that the relative neglect of apparent competition in

the past has been unwarranted, but also re-emphasizes that the distinction is important within interspecific competition between pattern on the one hand, and process or mechanism on the other. In the past, patterns of niche differentiation, and also of increased abundance of one species in the absence of another, have been interpreted as evidence of competition too readily. Now we can see that such patterns can arise through a wide variety of processes, and that a proper understanding requires that we distinguish between them – not only discriminating between

conventional and apparent competition, but also specifying mechanisms within, say, conventional competition (a point to which we return in Section 8.10).

8.7 Ecological effects of interspecific competition: experimental approaches

field and laboratory experiments

Notwithstanding the important interactions between competition and environmental heterogeneity, and the complications of apparent competition, a great deal of attention has been focused on conventional competition itself. We have already noted the difficulties in interpreting merely observational evidence (but see Freckleton & Watkinson, 2001), and it is for this reason that many studies of the ecological effects of interspecific competition have taken an experimental approach. For example, we have seen manipulative field experiments involving barnacles (see Section 8.2.2), birds (see Section 8.2.5), cattails (see Section 8.3.3) and snails (see Section 8.5.4), where the density of one or both species was altered (usually reduced). The fecundity, the survivorship, the abundance or the resource utilization of the remaining species was subsequently monitored. It was then compared either with the situation prior to the manipulation, or, far better, with a comparable control plot in which no manipulation had occurred. Such experiments have consistently provided valuable information, but they are typically easier to perform on some types of organism (e.g. sessile organisms) than they are on others.

The second type of experimental evidence has come from work carried out under artificial, controlled (often laboratory) conditions. Again, the crucial element has usually been a comparison between the responses of species living alone and their responses when in combination. Such experiments have the advantage of being comparatively easy to perform and control, but they have two major disadvantages. The first is that species are examined in environments that are different from those they experience naturally. The second is the simplicity of the environment: it may preclude niche differentiation because niche dimensions are missing that would otherwise be important. Nevertheless, these experiments can provide useful clues to the likely effects of competition in nature.

8.7.1 Longer term experiments

The most direct way of discovering the outcome of competition between two species in the laboratory, or under other controlled conditions, is to put them together and leave them to it. However, since even the most one-sided competition is likely to take a few generations (or a reasonable period of modular growth) before

it is completed, this direct approach is easier, and has been more frequently used, in some species than in others. It has most frequently been applied to insects (such as the flour beetle example in Section 8.4.3) and microorganisms (such as the *Paramecium* example in Section 8.2.4). Note that neither higher plants, nor vertebrates, nor large invertebrates, lend themselves readily to this approach (although a plant example is discussed in Section 8.10.1). We must be aware that this may bias our view of the nature of interspecific competition.

8.7.2 Single-generation experiments

Given these problems, the alternative 'laboratory' approach, especially with plants (although the methods have occasionally been used with animals), has generally been to follow populations over just a single generation, comparing 'inputs' and 'outputs'. A number of experimental designs have been used.

In 'substitutive' experiments, the effect of varying the proportion of each of two species is explored whilst keeping overall density constant (de Wit, 1960). Thus, at an overall density of say 200 plants, a series of mixtures would be set up: 100 of species A with 100 of species B, 150 A and 50 B, 0 A and 200 B, and so on. At the end of the experimental period, the amount of seed or the biomass of each species in each mixture would be monitored. Such replacement series may then be established at a range of total densities. In practice, however, most workers have used only a single total density, and this has led to considerable criticism of the design since it means that the effect of competition over several generations – when total density would inevitably alter – cannot be predicted (see Firbank & Watkinson, 1990).

substitutive experiments

None the less, replacement series have provided valuable insights into the nature of interspecific competition and the factors influencing its intensity (Firbank & Watkinson, 1990). An early, influential study was that of de Wit *et al.* (1966) on competition between the grass *Panicum maximum* and the legume *Glycine javanica*, which often form mixtures in Australian pastures. *Panicum* acquires its nitrogen only from the soil, but *Glycine* acquires part of its nitrogen from the air, by nitrogen fixation, through its root association with the bacterium *Rhizobium* (see Section 13.10.1). The competitors were grown in replacement series with and without an inoculation of *Rhizobium*, and the results are given both as replacement diagrams and as 'relative yield totals' (Figure 8.18). The relative yield of a species in a mixture is the ratio of its yield in the mixture to its yield alone in the replacement series, removing any absolute yield differences between species and referring both to the same scale. The relative yield total of a particular mixture is then the sum of the two relative yields. It is fairly clear from the replacement series (Figure 8.18a)

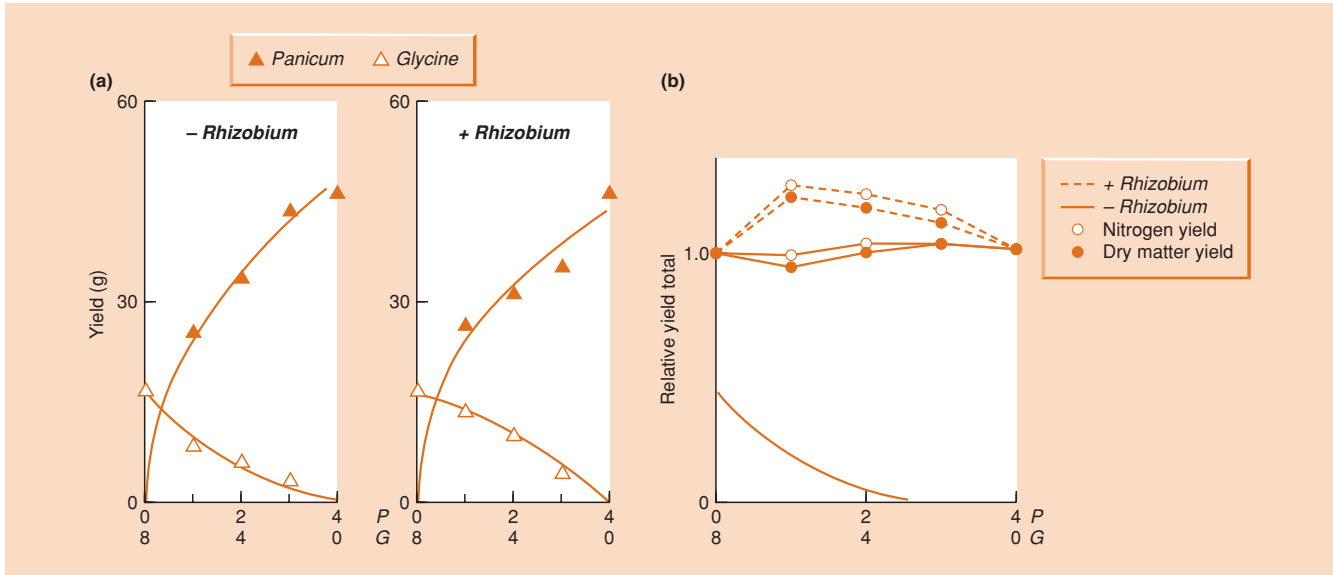


Figure 8.18 A substitutive experiment on interspecific competition between *Panicum maximum* (P), and *Glycine javanica* (G), in the presence and absence of *Rhizobium*: (a) replacement diagrams; (b) relative yield totals. (After de Wit *et al.*, 1966.)

that both species, but especially *Glycine*, fared better (were less affected by interspecific competition) in the presence than in the absence of *Rhizobium*. This is clearer still, however, from the relative yield totals (Figure 8.18b), which never departed significantly

from 1 in the absence of *Rhizobium*, but consistently exceeded 1 in its presence. This suggested that niche differentiation was not possible without *Rhizobium* (a second species could only be accommodated by a compensatory reduction in the output of the first) and that niche differentiation occurred in its presence (the species yielded more between them than either could alone).

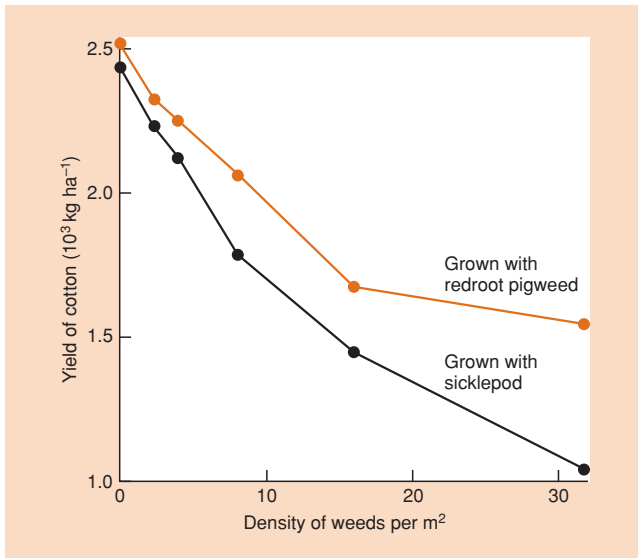


Figure 8.19 An 'additive design' competition experiment: the yield of cotton produced from stands planted at constant density, infested with weeds (either sicklepod or redroot pigweed) at a range of densities. (After Buchanan *et al.*, 1980.)

A second popular approach in the past has been the use of an 'additive' design, in which one species (typically a crop) is sown at a constant density, along with a range of densities of a second species (typically a weed). The justification for this is that it mimics the natural situation of a crop infested by a weed, and it therefore provides information on the likely effect on the crop of various levels of infestation (Firbank & Watkinson, 1990). A problem with additive experiments, however, is that overall density and species proportion are changed simultaneously. It has therefore proved difficult to separate the effect of the weed itself on crop yield from the simple effect of increasing total density (crop plus weed). An example is shown in Figure 8.19, describing the effects of two weeds, sicklepod (*Cassia obtusifolia*) and redroot pigweed (*Amaranthus retroflexus*), on the yield of cotton grown in Alabama (Buchanan *et al.*, 1980). As weed density increased, so cotton yield decreased, and this effect of interspecific competition was always more pronounced with sicklepod than with redroot pigweed.

In substitutive designs the proportions of competitors are varied but total density is held constant, whilst in

additive experiments

response surface analysis

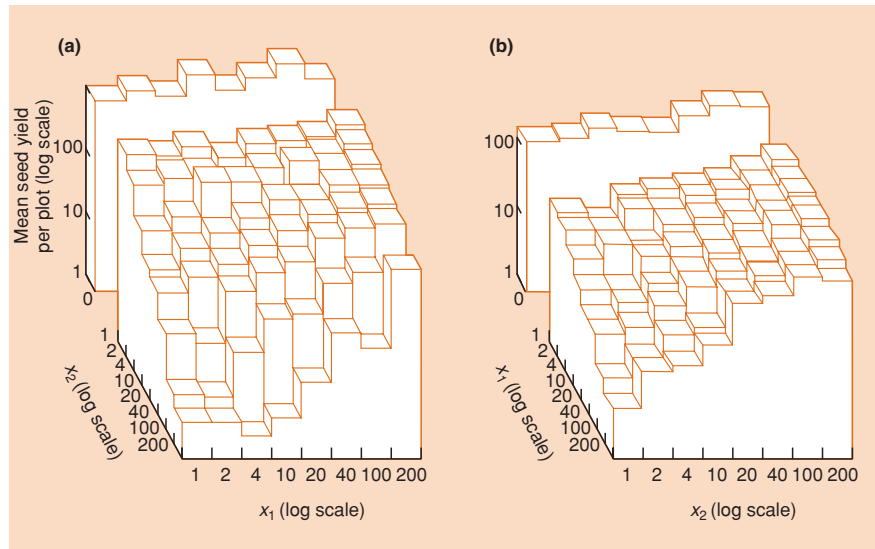


Figure 8.20 The response surface of competition, as indicated by seed production per pot, between (a) *Phleum arenarium* and (b) *Vulpia fasciculata* sown alone and in mixtures over a range of densities and frequencies. x_1 and x_2 are the sowing densities of Phleum and Vulpia, respectively. (After Law & Watkinson, 1987.)

additive designs the proportions are varied but the density of one competitor is held constant. It is perhaps not surprising, therefore, and certainly welcome, that a ‘response surface analysis’ has been proposed and applied, in which two species are grown alone and in mixtures at a wide range of densities and proportions (Figure 8.20) (Firbank & Watkinson, 1985; Law & Watkinson, 1987; Bullock *et al.*, 1994b; although the last of these deals with clones of the same species). Overall, these studies suggest that good equations for describing the competitive effect of one species (A) on another (B) are, for mortality:

$$N_A = N_{i,A} [1 + m(N_{i,A} + \beta N_{i,B})]^{-1} \quad (8.16)$$

and for fecundity:

$$Y_A = N_A R_A [1 + a(N_A + \alpha N_B)]^{-b}, \quad (8.17)$$

which can both be seen to be related to Equation 5.17 (see Section 5.8.1 – basic model of intraspecific competition) and Equation 5.12 (see Section 8.4.1 – incorporation of interspecific competition). Thus, $N_{i,A}$ and $N_{i,B}$ are the initial numbers of species A and B; N_A and N_B are the numbers of species A and B after mortality; Y_A is the yield (seeds or biomass) of species A; m and a are susceptibilities to crowding; β and α are competition coefficients; R_A is the basic reproductive rate of species A (and hence, $N_A R_A$ is the yield in the absence of competition); and b determines the type of density dependence (assumed equal to 1 for mortality – perfect compensation). Data like those shown in Figure 8.20, obtained over a single generation, can thus be used to fit values (by computer program) to the parameters in Equations 8.16 and 8.17, and the equations in their turn can be used to predict the

outcome of competition between the species over many generations – which is not possible with either substitutive or additive designs.

On the other hand, Law and Watkinson (1987) found that they could obtain an improved fit to their response surfaces, especially for one of their species, if they used an equation in which competition coefficients were not fixed, but varied with both frequency and density – although the meaning of this in terms of ‘plant behavior’ is not clear. Hence, response surface analyses, in revealing the potential complexities in interactions between competing species, also reveal that knowing or predicting dynamical outcomes may be only part of the story. It may also be necessary to understand the underlying mechanisms (see Section 8.10).

8.8 Evolutionary effects of interspecific competition

8.8.1 Natural experiments

We have seen that interspecific competition is commonly studied by an experimenter comparing species alone

pros and cons of natural experiments

and in combination. Nature, too, often provides information of this sort: the distribution of certain potentially competing species is such that they sometimes occur together (sympatry) and sometimes occur alone (allopatriy). These ‘natural experiments’ can provide additional information about interspecific competition, and especially about evolutionary effects, since the differences between sympatric and allopatric populations are often of long standing. The attractions of natural experiments are first that they

are natural – they are concerned with organisms living in their natural habitats – and second, that they can be ‘carried out’ simply by observation – no difficult or impracticable experimental manipulations are required. They have the disadvantage, however, of lacking truly ‘experimental’ and ‘control’ populations. Ideally, there should be only one difference between the populations: the presence or absence of a competitor species. In practice, though, populations typically differ in other ways too, simply because they exist in different locations. Natural experiments should therefore always be interpreted cautiously.

competitive release and character displacement

Evidence for competition from natural experiments usually comes either from niche expansion in the absence of a competitor (known as *competitive release*) or simply from a difference in

the realized niche of a species between sympatric and allopatric populations. If this difference is accompanied by morphological changes, then the effect is referred to as *character displacement*. On the other hand, physiological, behavioral and morphological traits are all equally likely to be involved in competitive interactions and to be reflections of a species’ realized niche. One difference may be that morphological distinctions are most obviously the result of evolutionary change, but as we shall see, physiological and behavioral ‘characters’ are also liable to ‘competitive displacement’.

gerbils in Israel: competitive release

One example of natural competitive release is provided by work on two gerbilline rodents living in the coastal sand dunes of Israel (Abramsky

& Sellah, 1982). In northern Israel, the protrusion of the Mt Carmel ridge towards the sea separates the narrow coastal strip into two isolated areas, north and south. *Meriones tristrami* is a gerbil that has colonized Israel from the north. It now occurs, associated with the dunes, throughout the length of the coast, including the areas both north and south of Mt Carmel. *Gerbillus allenbyi* is another gerbil, also associated with the dunes and feeding on similar seeds to *M. tristrami*; but this species has colonized Israel from the south and has not crossed the Mt Carmel ridge. To the north of Mt Carmel, where *M. tristrami* lives alone, it is found on sand as well as other soil types. However, south of Mt Carmel it occupies several soil types but not the coastal sand dunes. Here, only *G. allenbyi* occurs on dunes.

This appears to be a case of competitive exclusion and competitive release: exclusion of *M. tristrami* by *G. allenbyi* from the sand to the south of Mt Carmel; release of *M. tristrami* to the north. Is this present day competitive exclusion, however, or an evolutionary effect? Abramsky and Sellah set up a number of plots south of Mt Carmel from which *G. allenbyi* was removed, and they compared the densities of *M. tristrami* in these plots with those in a number of similar control plots. They monitored the plots for 1 year, but the abundance of *M. tristrami* remained essentially unchanged. It seems that south of Mt Carmel, *M. tristrami* has

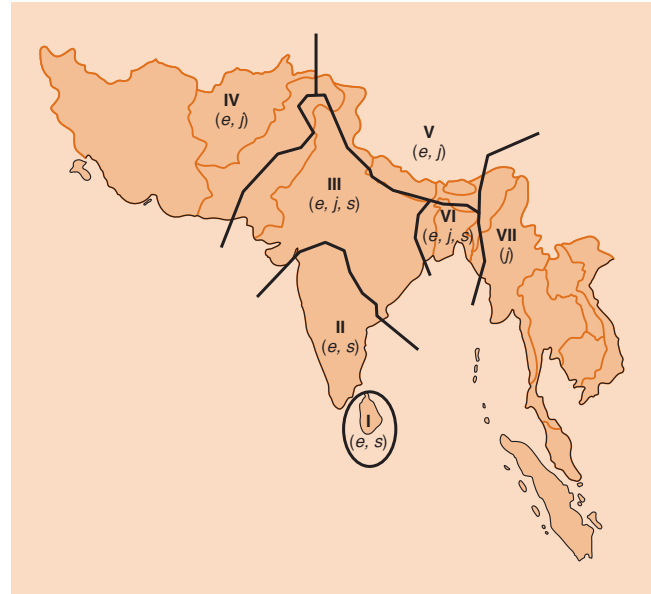


Figure 8.21 Native geographic ranges (I–VII) of *Herpestes javanicus* (j), *H. edwardsii* (e) and *H. smithii* (s). (From Simberloff *et al.*, 2000.)

evolved to select those habitats in which it avoids competition with *G. allenbyi*, and that even in the absence of *G. allenbyi* it retains this genetically fixed preference. Note, though, as ever, that this interpretation, because it invokes the ghost of competition past, may be sound and sensible – but it is not established fact.

A case of apparent morphological character displacement comes from work on Indian mongooses. In the western parts of its natural range, the small Indian mongoose (*Herpestes javanicus*) coexists with one or two slightly larger species in the same genus (*H. edwardsii* and *H. smithii*), but these species are absent in the eastern part of its range (Figure 8.21). Simberloff *et al.* (2000) examined size variation in the upper canine tooth, the animal’s principal prey-killing organ (note that female mongooses are smaller than males). In the east where it occurs alone (area VII in Figure 8.21), both males and females have larger canines than in the western areas (III, V, VI) where it coexists with the larger species (Figure 8.22). This is consistent with the view that where similar but larger predators are present, the prey-catching apparatus of *H. javanicus* has been selected for reduced size. This is likely to reduce the strength of competition with other species in the genus because smaller predators tend to take smaller prey than larger predators. Where *H. javanicus* occurs in isolation, its canine teeth are much larger.

morphological character displacement . . . in Indian mongooses . . .

It is of particular interest that the small Indian mongoose was introduced about a century ago to many islands outside its

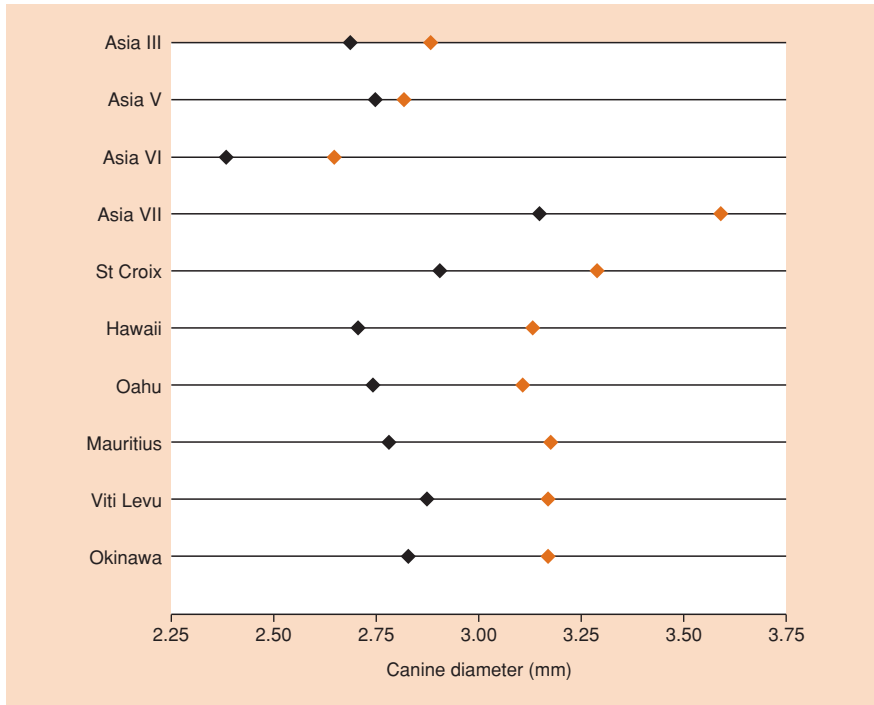


Figure 8.22 Maximum diameter (mm) of the upper canine for *Herpestes javanicus* in its native range (data only for areas III, V, VI and VII from Figure 8.21) and introduced range. Black symbols represent mean female size and colored symbols represent mean male size. (From Simberloff *et al.*, 2000.)

native range (often as part of a naive attempt to control introduced rodents). In these places, the larger competitor mongoose species are absent. Within 100–200 generations the small Indian mongoose has increased in size (Figure 8.22), so that the sizes of island individuals are now intermediate between those in the region of origin (where they coexisted with other species and were small) and those in the east where they occur alone. On the islands they show variation consistent with ‘ecological release’ from competition with larger species.

... and in three-spined sticklebacks in Canada

A further example concerns populations of the originally marine three-spined stickleback, *Gasterosteus aculeatus*, living in freshwater lakes in British Columbia, Canada, having apparently been left behind either following uplifting of the land after deglaciation, around 12,500 years ago, or after the subsequent rise and fall of sea levels around 11,000 years ago (Schluter & McPhail, 1992, 1993). As a result of this ‘double invasion’, some lakes now support two species of *G. aculeatus* (although they have not, as yet, been given their own specific names), whilst others support only one. Wherever there are two species, one is always ‘limnetic’, the other ‘benthic’. The first concentrates its feeding on plankton in the open water and has correspondingly long (and closely spaced) gill rakers that sieve the plankton from the stream of ingested water. The second, with much shorter gill rakers, concentrates on larger prey that it consumes largely from vegetation or sediments (Figure 8.23b). Wherever there is only one species in a lake, however, this species exploits both food

resources and is morphologically intermediate (Figure 8.23a). Presumably, either ecological character displacement has evolved since the second invasion, and this has promoted the coexistence of the species pairs, or it was a necessary prerequisite for the second invasion to be successful. Genetic evidence, based on

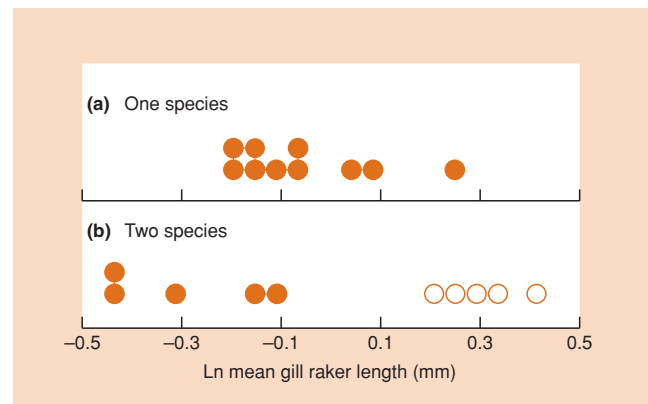


Figure 8.23 Character displacement in three-spined sticklebacks (*Gasterosteus aculeatus*). In small lakes in coastal British Columbia supporting two stickleback species (b), the gill rakers of the benthic species (●) are significantly shorter than those of the limnetic species (○), whilst those species of sticklebacks that occupy comparable lakes alone (a) are intermediate in length. Lengths of gill rakers have been adjusted to take account of species differences in overall size. (After Schluter & McPhail, 1993.)

analyses of mitochondrial DNA of several species pairs, supports the idea of repeated patterns of adaptive radiation within individual lakes (Rundle *et al.*, 2000).

If character displacement has ultimately been caused by competition, then the effects of competition should decline with the degree of displacement. Brook sticklebacks (*Culaea inconstans*) that are sympatric in Canadian lakes with ninespine sticklebacks (*Pungitius pungitius*) possess significantly shorter gill rakers, longer jaws and deeper bodies than allopatric brook sticklebacks. Gray and Robinson (2002) view allopatric brook sticklebacks as pre-displacement phenotypes and sympatric brook sticklebacks as postdisplacement phenotypes. When each phenotype was separately placed in enclosures in the presence of ninespine sticklebacks, the allopatric (pre-displacement) brook sticklebacks grew significantly less well than their sympatric (postdisplacement) counterparts (Figure 8.24). This is consistent with the hypothesis that competition is reduced when divergence between competing species occurs.

mud snails: a classic example of character displacement?

Two final, plausible examples of character displacement are provided by work on mud snails in Finland (*Hydrobia ulvae* and *H. ventrosa*) and giant rhinoceros beetles in Southeast

Asia (*Chalcosoma caucasus* and *C. atlas*). When the two mud snail species live apart, their sizes are more or less identical; but when they coexist they are always different in size (Figure 8.25a)

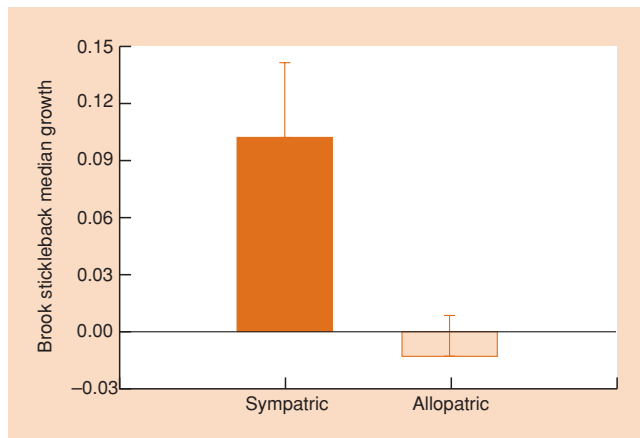


Figure 8.24 Means (with standard error) of group-median growth (natural log of the final mass of fish in each enclosure divided by the initial mass of the group) for sympatric brook sticklebacks representing postdisplacement phenotypes (dark orange bar) and allopatric brook sticklebacks representing pre-displacement phenotypes (light bar), both reared in the presence of ninespine sticklebacks. In competition with ninespine sticklebacks, growth was significantly greater for postdisplacement versus pre-displacement phenotypes ($P = 0.012$). (After Gray & Robinson, 2002.)

(Saloniemi, 1993) and they tend to consume different food particle sizes (Fenchel 1975). The beetles display a similar morphological pattern (Figure 8.25b) (Kawano, 2002). These data, therefore, strongly suggest character displacement, allowing coexistence. However, even an apparently exemplary example such as that of the mud snails is open to serious question (Saloniemi, 1993). In Finland, the sympatric and allopatric habitats were not identical: *H. ulvae* and *H. ventrosa* coexisted in sheltered water bodies rarely affected by tidal action, *H. ulvae* was found alone in relatively exposed tidal mudflats and salt marshes, and *H. ventrosa* was found alone in nontidal lagoons and pools. Moreover, *H. ulvae* naturally grows larger in less tidal habitats, and *H. ventrosa* may grow less well in this habitat. This alone could account for the size differences between sympatry and allopatry in these species. This emphasizes the major problem with natural experiments such as those that seem to demonstrate character displacement: sympatric and allopatric populations can occur in different environmental conditions over which the observer has no control. Sometimes it will be these environmental differences, rather than competition, that have led to the character displacement.

8.8.2 Experimenting with natural experiments

Sometimes, as we have already seen with the gerbils, natural experiments may themselves provide an opportunity for a further – and more informative –

niche divergence in clover–grass competition

experimental manipulation. In one such case, niche divergence was sought in clover, *Trifolium repens*, as a result of its having to compete with the grass *Lolium perenne* (Turkington & Mehrhoff, 1990). Clover was examined from two sites: (i) a ‘two-species’ site, in which clover achieved a ground coverage of 48% and the grass achieved a coverage of 96% (the two added together exceed 100% because their leaves can overlap); and (ii) a site in which clover achieved 40% coverage, but *L. perenne* covered only 4% (effectively a ‘clover-alone’ site). A total of three transplant (into the other site) and three re-plant (back into the home site) experiments were carried out (described and numbered in Figure 8.26a). *T. repens*, from both sites, was planted in: (i) plots at the two-species site cleared of *T. repens* only; (ii) plots at the two-species site cleared of both *T. repens* and *L. perenne*; and (iii) plots at the clover-alone site cleared of *T. repens*. The extent of competitive suppression or release was assessed from the amount of growth achieved by the different plantings of *T. repens*. From this the extent of the evolution of niche divergence between ‘clover-alone’ and ‘two-species’ *T. repens* was deduced, as was that between *T. repens* and *L. perenne*.

The *T. repens* population from the two-species site had indeed apparently diverged from the *L. perenne* population with which it was coexisting (and with which it may otherwise have competed strongly), and had diverged too from the clover-alone

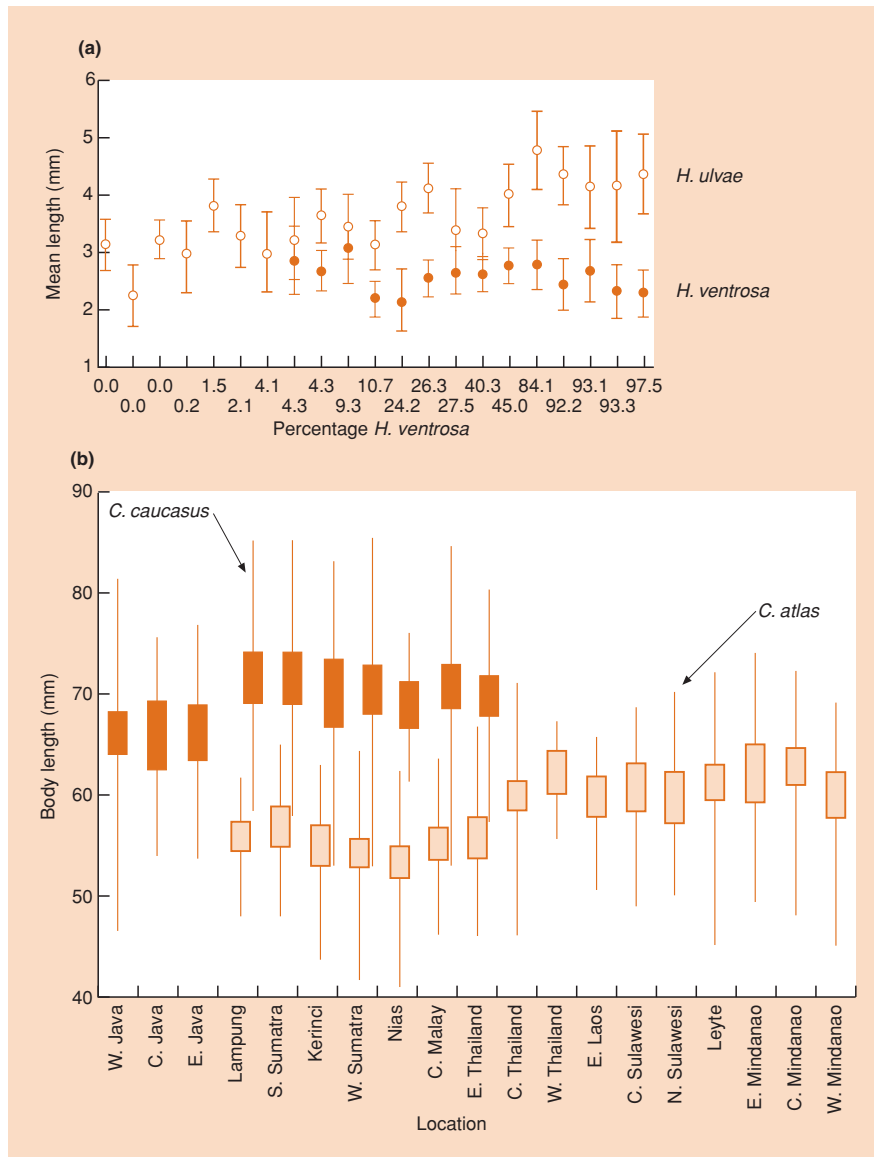


Figure 8.25 Character displacement in body size. (a) Mud snails in Finland (average lengths of *Hydrobia ulvae* and *H. ventrosa*, arranged in order of increasing percentage of *H. ventrosa*). (After Saloniemi, 1993.) (b) Giant rhinoceros beetles in southeast Asia (average lengths of *Chalcosoma caucasus* and *C. atlas*). (After Kawano, 2002.) In each case in allopatry, body sizes overlap broadly, but in sympatry body sizes are significantly different.

population (Figure 8.26b). When the two-species site was cleared of *T. repens* only, the re-planted *T. repens* grew better than the transplanted clover-alone plants (treatments 1 and 4, respectively; $P = 0.086$, close to significance), suggesting that the transplanted plants were competing more with the resident *L. perenne*. Moreover, when *L. perenne* was also removed, this made no difference to the two-species *T. repens* (treatments 4 and 5; $P > 0.9$), but led to a large increase in the growth of the clover-alone plants (treatments 1 and 2; $P < 0.005$). Also, when *L. perenne* was removed, the clover-alone plants grew better than the two-species ones (treatments 2 and 5; $P < 0.05$) – all of which suggests that only the clover-alone plants were released from competition by the absence of *L. perenne*. Finally, at the clover-alone site, the two-species clover plants grew no better than they had at their

home site (treatments 4 and 6; $P > 0.7$), whereas the clover-alone plants grew far better than they had at the two-species site in the presence of the grass (treatments 1 and 3; $P < 0.05$). Thus, the clover from the two-species population hardly competes with the *L. perenne* with which it coexists, whereas the clover-alone population would do – and does so if transplanted to the two-species site.

8.8.3 Selection experiments

The most direct way of demonstrating the evolutionary effects of competition within a pair of competing species is for the experimenter to induce these

direct demonstrations of evolutionary effects of competition have been rare

Figure 8.26 (a) Experimental design to test for the evolution of *Trifolium repens* (T) in competition with *Lolium perenne* (L). Indigenous populations of *T. repens*, and sometimes also *L. perenne*, were removed. *Trifolium repens* was removed from the base of the arrow and transplanted, or replanted, at the head of the arrow. Treatment numbers are consistent with the usage of Connell (1980). (b) The results of this experiment are in terms of the total plot dry weight achieved by *T. repens* in the various treatments. Significance levels for comparisons between pairs of treatments are given in the text. (After Turkington & Mehrhoff, 1990.)

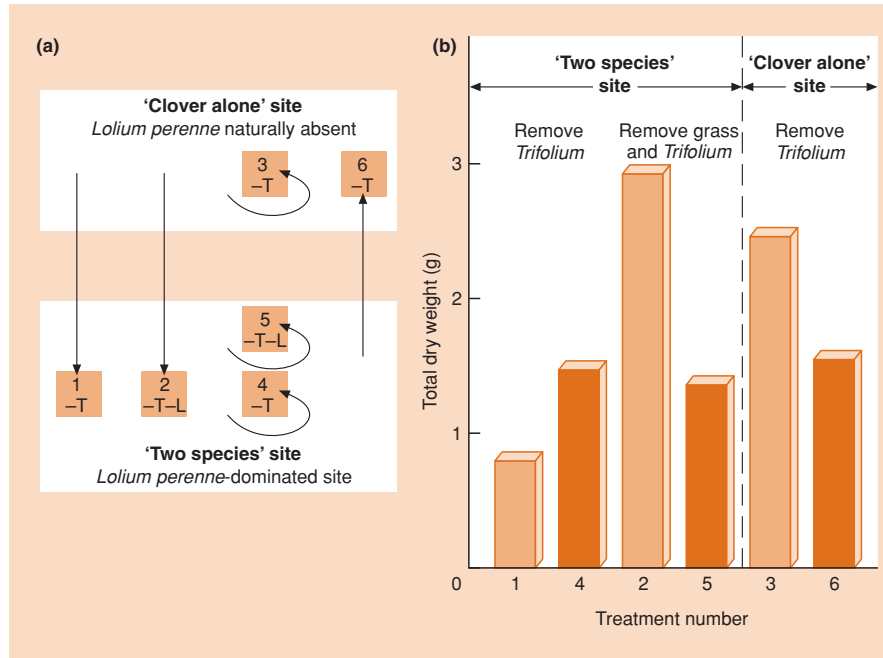
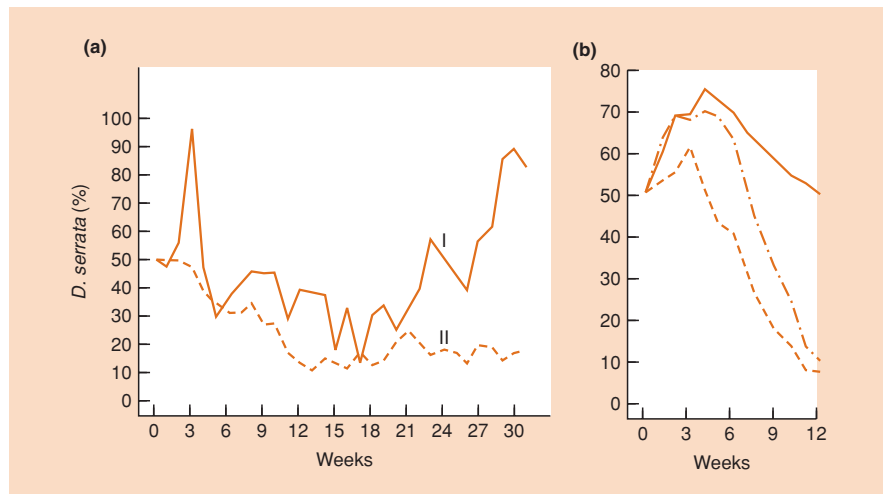


Figure 8.27 Apparent evolution of competitive ability in *Drosophila serrata*. (a) Of two experimental populations coexisting (and competing) with *D. nebulosa*, one (I) increased markedly in frequency after around week 20. (b) Individuals from this population did better in further competition with *D. nebulosa* ((—), mean of five populations) than did individuals from population II ((- - -), mean of five), or individuals from a stock not previously subjected to interspecific competition ((- · - · -), mean of five). (After Ayala, 1969.)



effects – to impose the selection pressure (competition) and observe the outcome. Surprisingly perhaps, there have been very few successful experiments of this type. In some cases, a species has responded to the selection pressure applied by a second, competitor species by apparently increasing its ‘competitive ability’, in the sense of increasing its frequency within a joint population. An example of this with two species of *Drosophila* is shown in Figure 8.27. Such results, however, tell us nothing about the means by which such apparent increases were achieved (e.g. whether it was as a result of niche differentiation).

To find an example of a demonstrable increase in niche differentiation giving rise to coexistence of competitors in a

selection experiment, we must turn away from interspecific competition in the strictest sense to competition between three types of the same bacterial species, *Pseudomonas fluorescens*, which behave as separate species because they reproduce asexually (Rainey & Trevisano, 1998). The three types are named ‘smooth’ (SM), ‘wrinkly spreader’ (WS) and ‘fuzzy spreader’ (FS) on the basis of the morphology of their colonies plated out on solid medium. In liquid medium, they also occupy quite different parts of the culture vessel (Figure 8.28a). In vessels that were continually shaken, so that no separate niches for the different types could be established, an initially pure culture of SM individuals retained its purity (Figure 8.28b). But in the absence

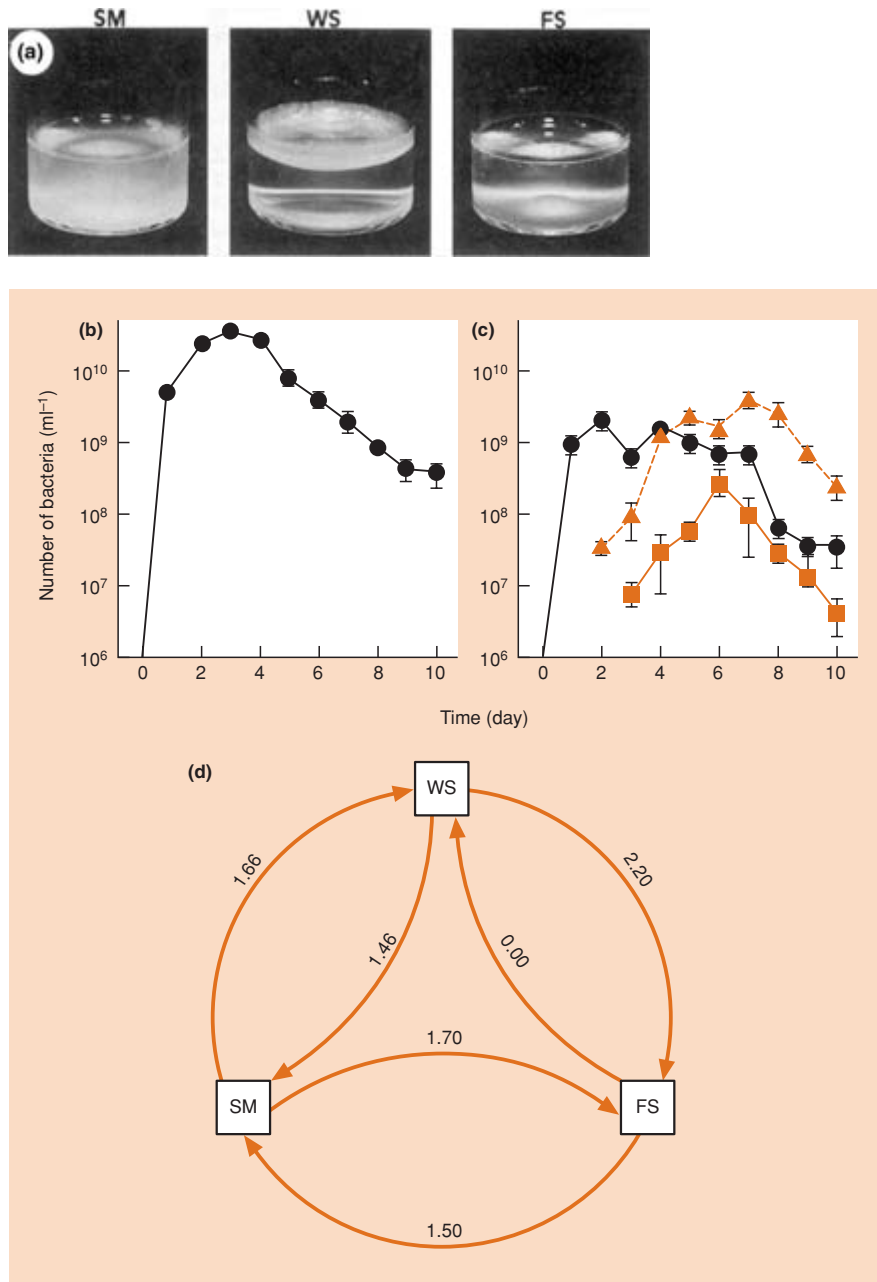


Figure 8.28 (a) Pure cultures of three types of the bacterium, *Pseudomonas fluorescens* (smooth, SM, wrinkly spreader, WS, and fuzzy spreader, FS) concentrate their growth in different parts of a liquid culture vessel. (b) In shaken culture vessels, pure SM cultures are maintained. Bars represent standard errors. (c) But in unshaken, initially pure SM cultures (\bullet), WS (\blacktriangle) and FS (\blacksquare) mutants arise, invade and establish. Bars represent standard errors. (d) The competitive abilities (relative rates of increase) when an initially rare type (foot of the arrow) invades a pure colony of another type (head of the arrow). Hence, values >1 indicate an ability to invade (superior competitor when rare) and values <1 an inability. (After Rainey & Trevisano, 1998. Reproduced by permission of Nature.)

of shaking, mutant WS and FS types invaded and established (Figure 8.28c). Furthermore, it was possible to determine the competitive abilities of the different types, when rare, to invade pure cultures of the other types (Figure 8.28d). Five of six possible invasions are favored. The exception – WS repels the invasion of FS – is unlikely to lead to the elimination of FS, because FS can invade cultures of SM, and SM can invade cultures of WS. In general, however, the experimental selection of increased niche differentiation amongst competing species appears to be either frustratingly elusive or sadly neglected.

8.9 Niche differentiation and similarity amongst coexisting competitors

It might be imagined that scientific progress is made by providing answers to questions. In fact, progress often consists of replacing one question with another, more pertinent, more challenging question. In this section, we deal with an area where this is the case: the questions of *how* different coexisting competitors are, and how different coexisting species need to be if competition is not to eliminate one of them.

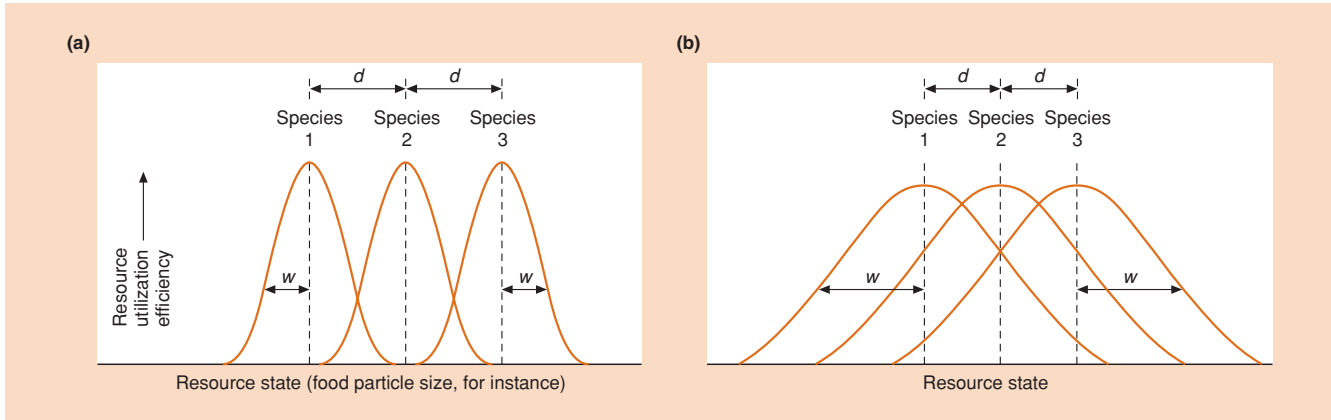


Figure 8.29 Resource-utilization curves for three species coexisting along a one-dimensional resource spectrum. d is the distance between adjacent curve peaks and w is the standard deviation of the curves. (a) Narrow niches with little overlap ($d > w$), i.e. relatively little interspecific competition. (b) Broader niches with greater overlap ($d < w$), i.e. relatively intense interspecific competition.

The Lotka–Volterra model predicts the stable coexistence of competitors in situations where interspecific competition is, for both species, less significant than intraspecific competition. Niche differentiation will obviously tend to concentrate competitive effects more within species than between them. The Lotka–Volterra model, and the Competitive Exclusion Principle, therefore imply that *any* amount of niche differentiation will allow the stable coexistence of competitors. Hence, in an attempt to discover whether this was ‘true’, the question ‘do competing species need to be different in order to coexist stably?’ greatly exercised the minds of ecologists during the 1940s (Kingsland, 1985).

how much niche differentiation is needed for coexistence?

It is easy to see now, however, that the question is badly put, since it leaves the precise meaning of ‘different’ undefined. We have seen examples in which the coexistence of competitors is apparently associated with some

degree of niche differentiation, but it seems that if we look closely enough, all coexisting species will be found to be different – without this having anything to do with competition. A more pertinent question, therefore, would be ‘is there a minimum amount of niche differentiation that has to be exceeded for stable coexistence?’ That is, is there a limit to the similarity of coexisting species?

One influential attempt to answer this question for exploitative competition, based on variants of the Lotka–Volterra model, was initiated by MacArthur and Levins (1967) and developed by May (1973). With hindsight, their approach is certainly open to question (Abrams, 1983). Nevertheless, we can learn most about the ‘limiting similarity problem’ by first examining their approach and then looking at the objections to it. Here, as so often, the models can be instructive without being ‘right’.

Imagine three species competing for a resource that is unidimensional and is distributed continuously; food size is a clear example. Each species has its own realized niche in this single dimension, which can be visualized as a resource-utilization curve (Figure 8.29). The consumption rate of each species is highest at the center of its niche and tails off to zero at either end, and the more the utilization curves of adjacent species overlap, the more the species compete. Indeed, by assuming that the curves are ‘normal’ distributions (in the statistical sense), and that the different species have similarly shaped curves, the competition coefficient (applicable to both adjacent species) can be expressed by the following formula:

a simple model provides a simple answer . . .

$$\alpha = e^{-d^2/4w^2} \tag{8.18}$$

where w is the standard deviation (or, roughly, ‘relative width’) of the curves, and d is the distance between the adjacent peaks. Thus, α is very small when there is considerable separation of adjacent curves ($d/w \gg 1$; Figure 8.29a), and approaches unity as the curves themselves approach one another ($d/w < 1$; Figure 8.29b).

How much overlap of adjacent utilization curves is compatible with stable coexistence? Assume that the two peripheral species have the same carrying capacity (K_1 , representing the suitability of the available resources for species 1 and 3) and consider the coexistence, in between them on the resource axis, of another species (carrying capacity K_2). When d/w is low (α is high and the species are similar) the conditions for coexistence are extremely restrictive in terms of the $K_1 : K_2$ ratio; but these restrictions lift rapidly as d/w approaches and exceeds unity (Figure 8.30). In other words, coexistence is possible when d/w

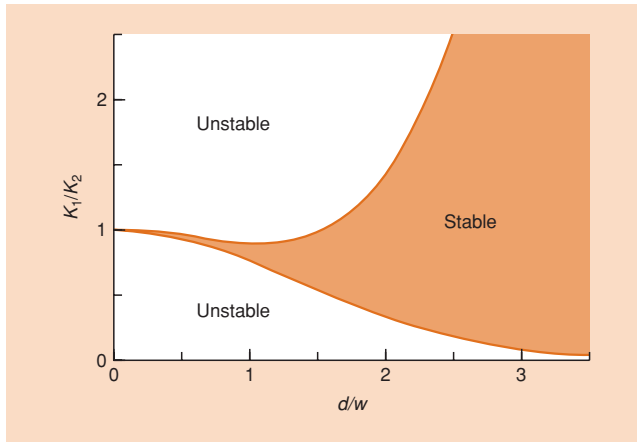


Figure 8.30 The range of habitat favorabilities (indicated by the carrying capacities K_1 and K_2 , where $K_1 = K_2$) that permit a three-species equilibrium community with various degrees of niche overlap (d/w). (After May, 1973.)

is low, but only if the suitabilities of the environment for the different species are extremely finely balanced. Furthermore, if the environment is assumed to vary, then the fluctuations will lead to variations in the $K_1 : K_2$ ratio, and coexistence will now only be possible if there is a broad range of $K_1 : K_2$ ratios leading to stability, i.e. if, roughly, $d/w > 1$.

... that is almost certainly wrong

This model, then, suggests that there is a limit to the similarity of coexisting competitors, and that the limit is given by the condition $d/w > 1$.

Are these the correct answers? In fact, it seems most unlikely that there is a universal limit to similarity, or even a widely applicable one that we could express in such a simple way as $d/w > 1$. Abrams (1976, 1983), amongst others, has emphasized that models with competition in several dimensions, with alternative utilization curves and so on, all lead to alternative limits to similarity, and often to much lower values of d/w being compatible with robust, stable coexistence. In other words, ' $d/w > 1$ ' is a property of one type of model analysis, but not of others, and thus, almost certainly, not of nature as a whole. Furthermore, we have already seen that because of environmental heterogeneity, apparent competition and so on, exploitative competition and any niche differentiation associated with it are not necessarily the whole story when it comes to the coexistence of competitors. This too argues against the idea of a universal limit.

On the other hand, the most general messages from the early models still seem valid, namely: (i) in the real world, with all its intrinsic variability, there are likely to be limits to the similarity of coexisting exploitative competitors; and (ii) these limits will reflect not only the differences between species, but also the variability within them, the nature of the resource, the nature of the utilization curve and so on.

But is even the limiting similarity question the best question to ask? We want to understand the extent of niche differentiation amongst coexisting species. If species were always packed as tightly together as they could be, then presumably they would differ by the minimum (limiting) amount. But why should they be? We return, once again, to the distinction between the ecological and the evolutionary consequences of competition (Abrams, 1990). The ecological effects are that species with 'inappropriate' niches will be eliminated (or repelled if they try to invade), and the limiting similarity question implicitly concerned itself with this: how many species can be 'packed in'? But coexisting competitors may also evolve. Do we generally observe the ecological effects, or the combined ecological and evolutionary effects? Do they differ? We cannot attempt to answer the first question without answering the second, and the answer to that seems to be, perhaps inevitably, 'it depends'.

the answer? it depends

Different models, based on different underlying mechanisms in the competitive process, predict that evolution will lead to more widely spaced niches, or to more closely packed niches or to much the same disposition of niches as those predicted by ecological processes alone (Abrams, 1990). Two points, therefore, emerge from this discussion. The first is that it has been entirely theoretical. This is a reflection of the second point, which is that we have seen progress – but in terms of successive questions superseding their predecessors rather than actually answering them. Data provide answers – what we have seen is a refinement of questions. The latest stage in this appears to be that attempts to answer questions regarding 'niche similarity' may need to be postponed until we know more about resource distributions, utilization curves and, more generally, the mechanisms underlying exploitative competition. It is to these that we now turn.

Two points, therefore, emerge from this discussion. The first is that it has been entirely theoretical. This is a reflection of the second point, which is that we have seen progress – but in terms of successive questions superseding their predecessors rather than actually answering them. Data provide answers – what we have seen is a refinement of questions. The latest stage in this appears to be that attempts to answer questions regarding 'niche similarity' may need to be postponed until we know more about resource distributions, utilization curves and, more generally, the mechanisms underlying exploitative competition. It is to these that we now turn.

8.10 Niche differentiation and mechanisms of exploitation

In spite of all the difficulties of making a direct connection between interspecific competition and niche differentiation, there is no doubt that niche differentiation is often the basis for the coexistence of species.

There are a number of ways in which niches can be differentiated. One is resource partitioning or, more generally, differential resource utilization. This can be observed when species living in precisely the same habitat nevertheless utilize different resources.

niche differentiation ...

... easy to imagine in animals, less easy in plants ...

Since the majority of resources for animals are individuals of other species (of which there are literally millions of types), or parts of individuals, there is no difficulty, in principle, in imagining how

competing animals might partition resources amongst themselves. Plants, on the other hand, all have very similar requirements for the same potentially limited resources (see Chapter 3), and there is much less apparent scope for resource partitioning (but see below).

In many cases, the resources used by ecologically similar species are separated spatially. Differential resource utilization will then express itself as either microhabitat differentiation between the species (e.g. different species of fish feeding at different depths), or even a difference in geographic distribution. Alternatively, the availability of the different resources may be separated in time; for example, different resources may become available at different times of the day or in different seasons. Differential resource utilization may then express itself as a temporal separation between the species.

... based on resources and conditions

The other major way in which niches can be differentiated is on the basis of conditions (Wilson, 1999). Two species may use precisely the same resources, but if their ability to do so is influenced by environmental conditions (as it is bound to be), and if they respond differently to those conditions, then each may be competitively superior in different environments. This too can express itself as a microhabitat differentiation, or a difference in geographic distribution or a temporal separation, depending on whether the appropriate conditions vary on a small spatial scale, a large spatial scale or over time. Of course, in a number of cases (especially with plants) it is not easy to distinguish between conditions and resources (see Chapter 3). Niches may then be differentiated on the basis of a factor (such as water) which is both a resource and a condition.

spatial and temporal separation

There are many examples of the separation of competing species in space or time involving both animals and plants. For example, tadpoles of two anuran species in New Jersey, USA (*Hyla crucifer* and *Bufo woodhousii*), have their feeding periods offset by around 4–6 weeks each year, apparently, though not certainly, associated with differential responses to environmental conditions rather than seasonal changes in resources (Lawler & Morin, 1993). Two coexisting species of spiny mice in rocky deserts in Israel partition activity on a diel basis: *Acomys cahirinus* is nocturnal and *A. russatus* is diurnal, although the latter becomes nocturnal if its congener is removed (Jones *et al.*, 2001). Two phloem-feeding bark beetles, *Ips duplicatus* and *I. typographus*, on Norway spruce trees, in Norway, are separated in their feeding sites on a small spatial scale by trunk diameter, although the reason for this is not at all clear (Schlyter & Anderbrandt, 1993). But, it is amongst plants and other sessile organisms, because of their limited scope for differential resource utilization at the same location and instant, that spatial and temporal separation are likely to be of particular significance (see Harper, 1977). Although, as ever, it is one thing to show that

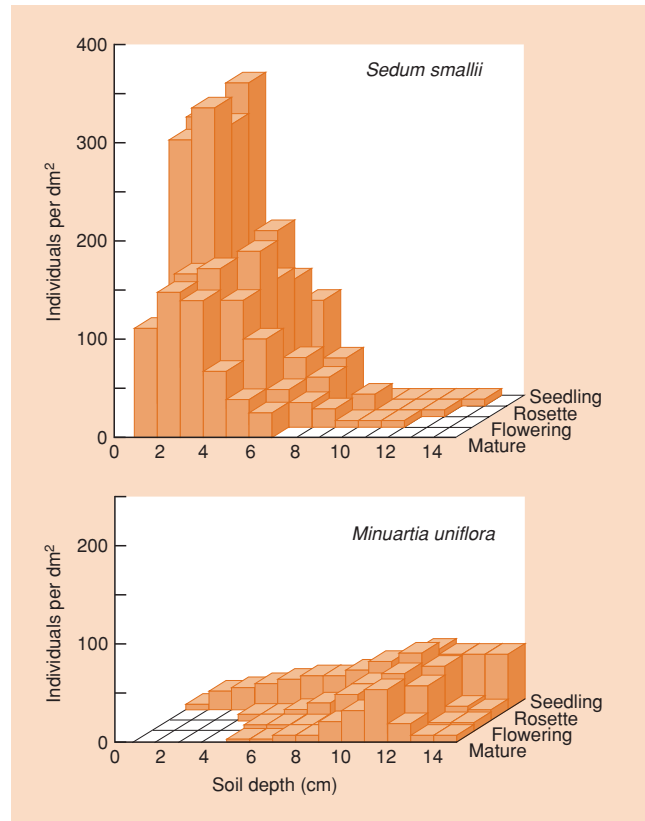


Figure 8.31 The zonation of individuals, according to soil depth, of two annual plants, *Sedum smallii* and *Minuartia uniflora* at four stages of the life cycle. (After Sharitz & McCormick, 1973.)

species differ in their spatial or temporal distribution – it is quite another to prove that this has anything to do with competition. The cattails in Section 8.3.3 provide one example of competing plants separated spatially. Another is shown in Figure 8.31, concerning the annuals *Sedum smallii* and *Minuartia uniflora* that dominate the vegetation growing on granite outcrops in south-eastern USA. The adult plants exhibit an especially clear spatial zonation associated with soil depth (itself strongly correlated with soil moisture), and further experimental results reinforce the idea that it is competition rather than mere differences in tolerance that gives rise to this zonation.

Describing the outcome of competition, however – ‘one species coexists with or excludes another’ – and even associating this with niche differentiation, whether based on resources themselves, or conditions or merely differences in space or time, actually provides us with rather little understanding of the competitive process. For this, as we have seen repeatedly in this chapter, we may need to focus more on the mechanisms of exploitation. *How*, precisely, does one species outexploit and outcompete another? *How* can two consumers coexist on two limiting resources, when both resources are absolutely essential to both consumers?

the need to consider resource dynamics

Furthermore, as Tilman (1990) has pointed out, whilst monitoring the population dynamics of two competing species may give us some powers of prediction for the next time they compete, it will give us very little help in predicting how each would fare against a third species. Whereas, if we understood the dynamics of the interaction of all the species with their shared limiting resources, then we might be able to predict the outcome of exploitative competition between any given pair of the species. We therefore turn now to some attempts to explain the coexistence of species competing for limiting resources that explicitly consider not only the dynamics of the competing species but also the dynamics of the resources themselves. Rather than going into details, we examine the outlines of models and some major conclusions.

8.10.1 Exploitation of a single resource

a model for a single resource . . .

Tilman (1990) shows, for a number of models, what we have already seen demonstrated empirically in Section 8.2.6, that when two species compete exploitatively for a single limiting resource the outcome is determined by which species is able, in its exploitation, to reduce the resource to the lower equilibrium concentration, R^* . (Satisfyingly, for apparent competition the reverse is true: the prey or host able to support the greatest abundance of predators or parasites is the winner (see, for example, Begon & Bowers, 1995) – a prediction we have seen borne out in Figure 8.17.)

Different models, based on varying details in the mechanism of exploitation, give rise to different formulae for R^* , but even the simplest model is revealing, giving:

$$R_i^* = m_i C_i / (g_i - m_i). \quad (8.19)$$

Here m_i is the mortality or loss rate of consumer species i ; C_i is the resource concentration at which species i attains a rate of growth and reproduction per unit biomass (relative rate of increase, RRI) equal to half its maximal RRI (C_i is thus highest in consumers that require the most resource in order to grow rapidly); and g_i is the maximum RRI achievable by species i . This suggests that successful exploitative competitors (low R_i^*) are those that combine resource-utilization efficiency (low C_i), low rates of loss (low m_i) and high rates of increase (high g_i). On the other hand, it may not be possible for an organism to combine, say, low C_i and high g_i . A plant's growth will be most enhanced by putting its matter and energy into leaves and photosynthesis – but to enhance its nutrient-utilization efficiency it would have to put these into roots. A lioness will be best able to subsist at low densities of prey by being fleet-footed and maneuverable – but this may be difficult if she is often heavily pregnant. Understanding successful exploitative competitiveness, therefore, may require us ultimately to

understand how organisms trade off features giving rise to low values of R^* against features that enhance other aspects of fitness.

A rare test of these ideas is provided by Tilman's own work on terrestrial plants competing for nitrogen (Tilman & Wedin, 1991a, 1991b).

. . . tested on grasses

Five grass species were grown alone in a range of experimental conditions that gave rise in turn to a range of nitrogen concentrations. Two species, *Schizachyrium scoparium* and *Andropogon gerardi*, consistently reduced the nitrate and ammonium concentrations in soil solutions to lower values than those achieved by the other three species (in all soils but those with the very highest nitrogen levels). Of these three other species, one, *Agrostis scabra*, left behind higher concentrations than the other two, *Agropyron repens* and *Poa pratensis*. Then, when *A. scabra* was grown with *A. repens*, *S. scoparium* and *A. gerardi*, the results, especially at low nitrogen concentrations where nitrogen was most likely to be limiting, were very much in line with the exploitative competition theory (Figure 8.32). The species that could reduce nitrogen to the lowest concentration always won – *A. scabra* was always competitively displaced. A similar result has been obtained for the nocturnal, insectivorous gecko *Hemidactylus frenatus*, an invader of urban habitats across the Pacific basin, where it is responsible for population declines of the native gecko *Lepidodactylus lugubris*. Petren and Case (1996) established that insects are a limiting resource for both. The invader is capable of depleting insect resources in experimental enclosures to lower levels than the native gecko, and the latter suffers reductions in body condition, fecundity and survivorship as a result.

Returning to Tilman's grasses, the five species were chosen from various points in a typical old-field successional sequence in Minnesota (Figure 8.33a), and it is clear that the better competitors for nitrogen are found later in the sequence. These species, and *S. scoparium* and *A. gerardi* in particular, had higher root allocations, but lower above-ground vegetative growth rates and reproductive allocations (e.g. Figure 8.33b). In other words, they achieved their low values of R^* by the high resource-utilization efficiency given to them by their roots (low C_i , Equation 8.19), even though they appeared to have paid for this through a reduction in growth and reproductive rates (lower g_i). In fact, over all the species, a full 73% of the variance in the eventual soil nitrate concentration was explained by variations in root mass (Tilman & Wedin, 1991a). This successional sequence (see Section 16.4 for a much fuller discussion of succession) therefore appears to be one in which fast growers and reproducers are replaced by efficient and powerful exploiters and competitors.

8.10.2 Exploitation of two resources

Tilman (1982, 1986; see also Section 3.8) has also considered what happens when two competitors compete for two resources. Beginning with *intraspecific*

a model for two resources – the zero net isocline: a niche boundary

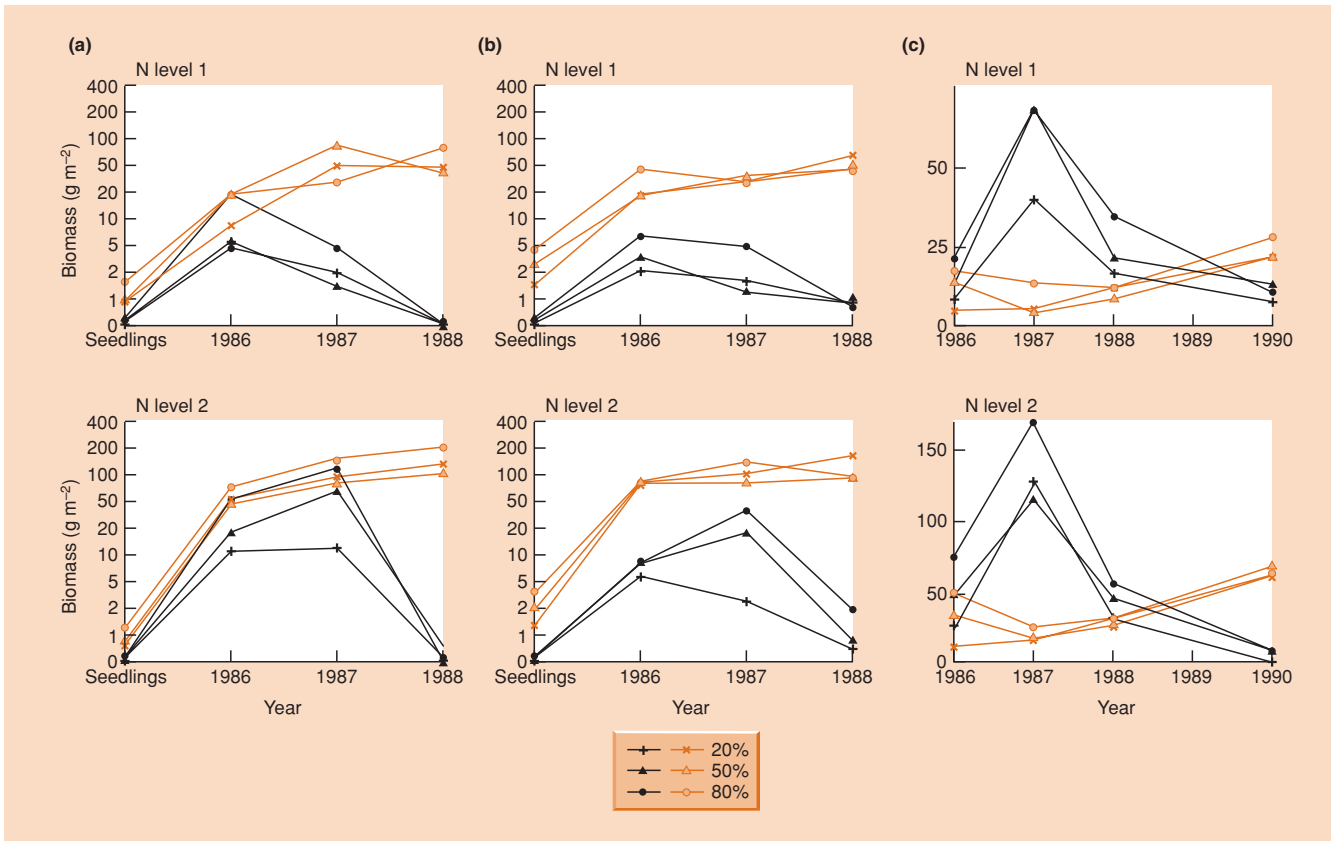


Figure 8.32 The results of competition experiments in which *Agrostis scabra* (black lines) was competitively displaced by (a) *Schizachyrium scoparium*, (b) *Andropogon gerardi* and (c) *Agropyron repens* (orange lines, at each of two nitrogen levels (both low) and whether it represented 20, 50 or 80% of the initial seed sown). In each case, *A. scabra* had lower values of R^* for nitrate and ammonium (see text). Displacement was least rapid in (c) where the differential was least marked. (After Tilman & Wedin, 1991b.)

competition, we can define a zero net growth isocline for a single species utilizing two essential resources (see Section 3.8). This isocline is the boundary between resource combinations that allow the species to survive and reproduce, and resource combinations that do not (Figure 8.34), and therefore represents the boundary of the species' niche in these two dimensions. For present purposes, we can ignore the complications of overcompensation, chaos, etc. and assume that intraspecific competition brings the population to a stable equilibrium. Here, however, the equilibrium has two components: both population size and the resource levels should remain constant. Population size is constant (by definition) at all points on the isocline, and Tilman established that there is only one point on the isocline where resource levels are also constant (point S^* in Figure 8.34). This point, which is the two-resource equivalent of R^* for one resource, represents a balance between the consumption of the resources by the consumer (taking the resource concentrations towards the bottom left of the figure) and the natural renewal of the resources (taking the concentrations towards the top right). Indeed, in the absence of the consumer, resource renewal would take the resource concentrations to the 'supply point', shown in the figure.

To move from intra- to interspecific competition, it is necessary to superimpose the isoclines of two species on the same diagram (Figure 8.35). The two species will presumably have different consumption rates, but there will still be a single supply point. The outcome depends on the position of this supply point.

In Figure 8.35a, the isocline of species A is closer to both axes than that of species B. There are three regions in which the supply point might be found. If it was in region 1, below the isoclines of both species, then there would never be sufficient resources for either species and neither would survive. If it was in region 2, below the isocline of species B but above that of species A, then species B would be unable to survive and the system would equilibrate on the isocline of species A. If the supply point was in region 3, then this system too would equilibrate on the isocline of species A. Analogous to the one-resource case, species A would competitively exclude species B because of its ability to exploit both resources down to levels at which species B could not survive. Of course, the outcome would be reversed if the positions of the isoclines were reversed.

a superior and an inferior competitor

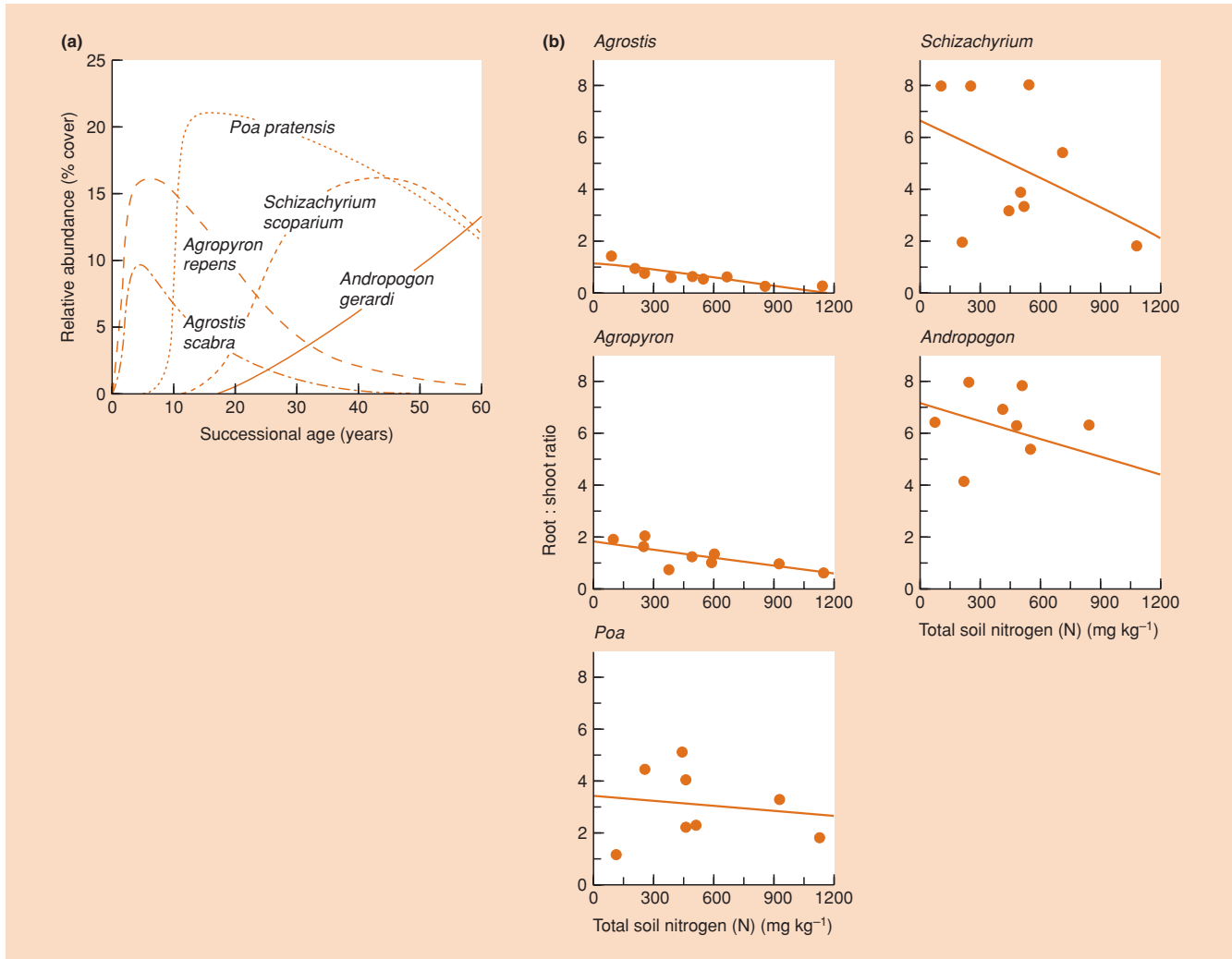


Figure 8.33 (a) The relative abundances of five grasses during old-field successions at Cedar Creek Natural History Area, Minnesota, USA. (b) The root : shoot ratios were generally higher in the later successional species and declined as soil nitrogen increased. (After Tilman & Wedin, 1991a.)

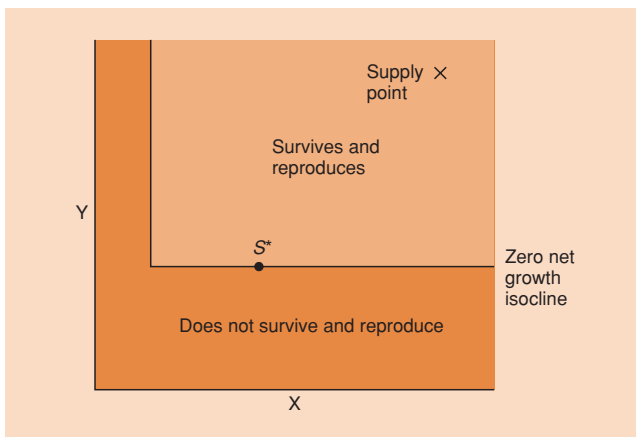


Figure 8.34 (left) The zero net growth isocline of a species potentially limited by two resources (X and Y), divides resource combinations on which the species can survive and reproduce, from those on which it cannot. The isocline is rectangular in this case because X and Y are essential resources (see Section 3.8.1). Point S* is the only point on the isocline at which there is also no net change in resource concentrations (consumption and resource renewal are equal and opposite). In the absence of the consumer, resource renewal would take the resource concentrations to the 'supply point' shown.

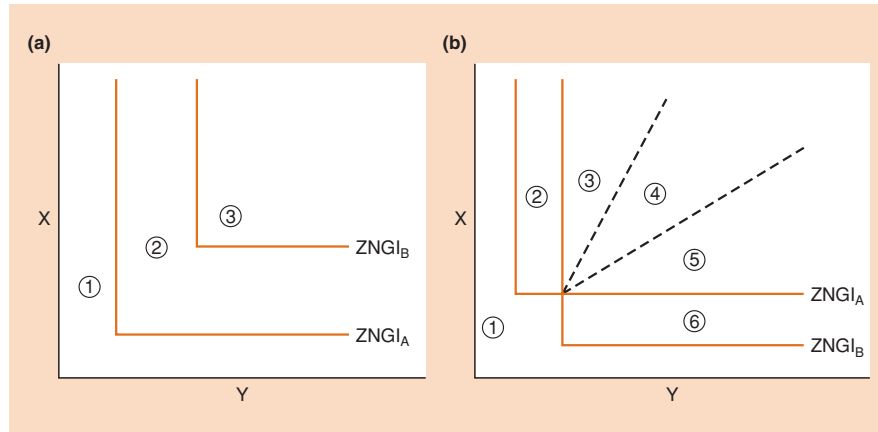


Figure 8.35 (a) Competitive exclusion: the isocline (zero net growth isocline, ZNGI) of species A lies closer to the resource axes than the isocline of species B. If the resource supply point is in region 1, then neither species can exist. But if the resource supply point is in regions 2 or 3, then species A reduces the resource concentrations to a point on its own isocline (where species B cannot survive and reproduce): species A excludes species B. (b) Potential coexistence of two competitors limited by two essential resources. The isoclines of species A and B overlap, leading to six regions of interest. With supply points in region 1 neither species can exist; with points in regions 2 and 3, species A excludes species B; and with points in regions 5 and 6, species B excludes species A. Region 4 contains supply points lying between the limits defined by the two dashed lines. With supply points in region 4 the two species coexist. For further discussion, see text.

coexistence – dependent on the ratio of resource levels at the supply point

In Figure 8.35b the isoclines of the two species overlap, and there are six regions in which the supply point might be found. Points in region 1 are below both isoclines and would allow neither species to exist; those in region 2 are below the isocline of species B and

would only allow species A to exist; and those in region 6 are below the isocline of species A and would only allow species B to exist. Regions 3, 4 and 5 lie within the fundamental niches of both species. However, the outcome of competition depends on which of these regions the supply point is located in.

The most crucial region in Figure 8.35b is region 4. For supply points here, the resource levels are such that species A is more limited by resource X than by resource Y, whilst species B is more limited by Y than X. However, species A consumes more X than Y, whilst species B consumes more Y than X. Because each species consumes more of the resource that more limits its own growth, the system equilibrates at the intersection of the two isoclines, and this equilibrium is stable: the species coexist.

subtle niche differentiation – each species consumes more of the resource that more limits its own growth

This is niche differentiation, but of a subtle kind. Rather than the two species exploiting different resources, species A disproportionately limits itself by its exploitation of resource X, whilst species B disproportionately limits itself by its exploitation of resource Y. The result is the coexistence of competi-

tors. By contrast, for supply points in region 3, both species are more limited by Y than X. But species A can reduce the level of

Y to a point on its own isocline below species B’s isocline, where species B cannot exist. Conversely, for supply points in region 5, both species are more limited by X than Y, but species B depresses X to a point below species A’s isocline. Thus, in regions 3 and 5, the supply of resources favors one species or the other, and there is competitive exclusion.

It seems then that two species can compete for two resources and coexist as long as two conditions are met.

- 1 The habitat (i.e. the supply point) must be such that one species is more limited by one resource, and the other species more limited by the other resource.
- 2 Each species must consume more of the resource that more limits its own growth. Thus, it is possible, in principle, to understand coexistence in competing plants on the basis of differential resource utilization. The key seems to be an explicit consideration of the dynamics of the resources as well as the dynamics of the competing species. As with other cases of coexistence by niche differentiation, the essence is that intraspecific competition is, for both species, a more powerful force than interspecific competition.

The best evidence for the validity of the model comes from Tilman’s own experimental laboratory work on competition between the diatoms *Asterionella formosa* and *Cyclotella meneghiniana* (Tilman, 1977). For both species, Tilman observed directly the consumption rates and the isoclines for both phosphate and silicate. He then used these to predict the outcome of competition with a range of resource supply points (Figure 8.35). Finally, he ran a number of competition experiments with a variety of supply

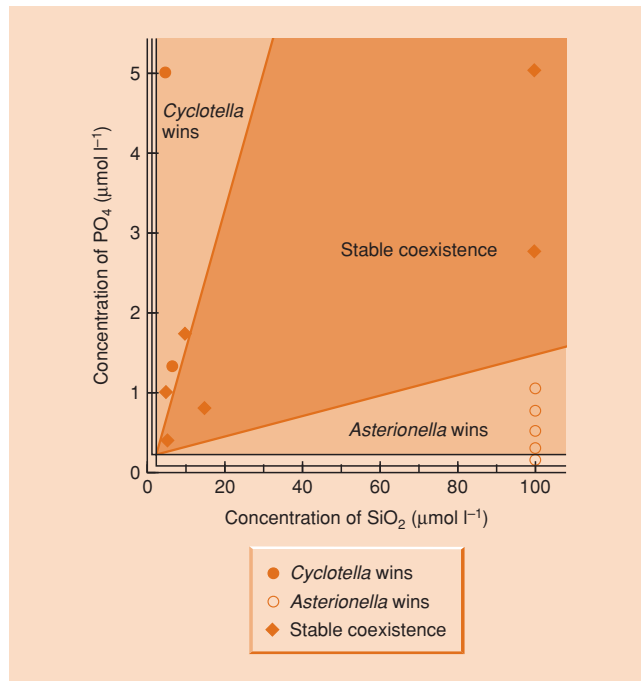


Figure 8.36 The observed isoclines and consumption vectors of two diatom species, *Asterionella formosa* and *Cyclotella meneghiniana*, were used to predict the outcome of competition between them for silicate and phosphate. The predictions were then tested in a series of experiments, the outcomes of which are depicted by the symbols explained in the key. Most experiments confirmed the predictions, with the exception of two lying close to the regional boundary. (After Tilman, 1977, 1982.)

points, and the results of these are illustrated in Figure 8.36. In most cases the results confirmed the predictions. In the two that did not, the supply points were very close to the regional boundary. The results are therefore encouraging. However, it has proved extremely difficult to transfer this approach from the laboratory, where supply points can be manipulated, to natural populations, where they cannot, and where even the estimation of supply points has proved practically impossible (Sommer, 1990). There is considerable need for consolidation and extension from work on other types of plants and animals.

8.10.3 Exploitation of more than two resources

the more limiting resources there are, the more species may coexist

We have seen how two diatom species may coexist in the laboratory on two shared limiting resources. In fact, Tilman's resource competition theory predicts that the diversity of coexisting species should be proportional to the

total number of resources in a system that are at physiological

limiting levels: the more limiting resources, the more coexisting competitors. Interlandi and Kilham (2001) tested this hypothesis directly in three lakes in the Yellowstone region of Wyoming, USA using an index (Simpson's index) of the species diversity of phytoplankton there (diatoms and other species). If one species exists on its own, the index equals 1; in a group of species where biomass is strongly dominated by a single species, the index will be close to 1; when two species exist at equal biomass, the index is 2; and so on. According to resource competition theory, this index should therefore increase in direct proportion to the number of resources limiting growth. The spatial and temporal patterns in phytoplankton diversity in the three lakes for 1996 and 1997 are shown in Figure 8.37. The principal limiting resources for phytoplankton growth are nitrogen, phosphorus, silicon and light. These parameters were measured at the same depths and times that the phytoplankton were sampled, and it was noted where and when any of the *potential* limiting factors *actually* occurred at levels below threshold limits for growth. Consistent with resource competition theory, species diversity increased as the number of resources at physiologically limiting levels increased (Figure 8.38).

These results suggest that even in the highly dynamic environments of lakes where equilibrium conditions are rare, resource competition plays a role in continuously structuring the phytoplankton community. It is heartening that the results of experiments performed in the artificial world of the laboratory are echoed here in the much more complex natural environment.

Our survey of interspecific competition has concluded with a realization that we need to understand much more about the mechanisms underlying the interactions between consumers and their resources. If these resources are alive, then we normally refer to such interactions as predation; and if they were alive once, but are now dead, we refer to them as detritivory. It would seem, therefore, that the distinction normally made between competition and predation is, in a very real sense, an artificial one (Tilman, 1990). None the less, having dealt with competition here, we turn next, in a separate series of chapters, to predation and detritivory.

Summary

In interspecific competition, individuals of one species suffer a reduction in fecundity, growth or survivorship as a result of resource exploitation or interference by individuals of another species. Competing species may exclude one another from particular habitats so that they do not coexist, or may coexist, perhaps by utilizing the habitat in slightly different ways. Interspecific competition is frequently highly asymmetric.

Although species may not be competing now, their ancestors may have done so in the past. Species may have evolved characteristics that ensure they compete less, or not at all, with other species. Moreover, species whose niches appear differentiated may have evolved independently and, in fact, never have

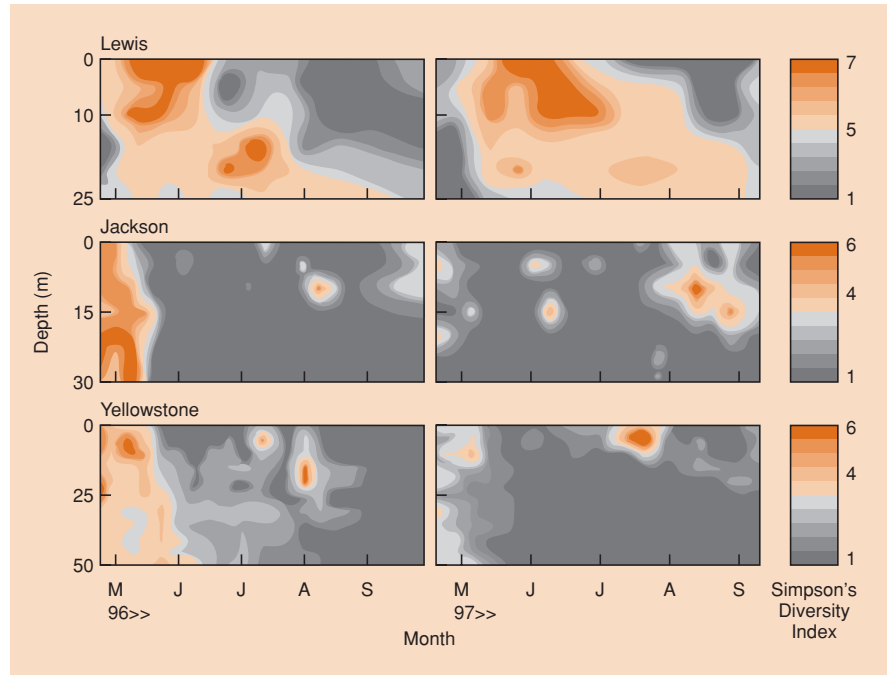


Figure 8.37 Variation in phytoplankton species diversity (Simpson's index) with depth in 2 years in three large lakes in the Yellowstone region, USA. Shading indicates depth-time variation in a total of 712 discrete samples: dark orange areas denote high species diversity, and gray areas denote low species diversity. (After Interlandi & Kilham, 2001.)

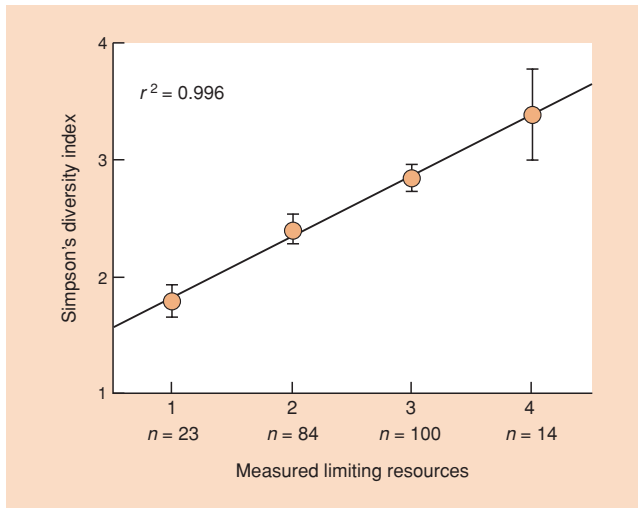


Figure 8.38 Phytoplankton diversity (Simpson's index; mean \pm SE) associated with samples with different numbers of measured limiting resources. It was possible to perform this analysis on 221 samples from those displayed in Figure 6.14. The number of samples (n) in each limiting resource class is shown. (From Interlandi & Kilham, 2001.)

competed, now or historically. An experimental manipulation (for instance, the removal of one or more species) can indicate the presence of current competition if it leads to an increase in the fecundity or survival or abundance of the remaining species.

But negative results would be equally compatible with the past elimination of species by competition, the evolutionary avoidance of competition in the past, and the independent evolution of noncompeting species.

Mathematical models, most notably the Lotka–Volterra model, have provided important insights into the circumstances that permit the coexistence of competitors, and those that lead to competitive exclusion. However, the simplified assumptions of the Lotka–Volterra model limit its applicability to real situations in nature. We know from other models and experiments that the outcome of interspecific competition can be strongly influenced by heterogeneous, inconstant or unpredictable environments. Coexistence of a superior and an inferior competitor on a patchy and ephemeral resource can occur if the two species have independent, aggregated distributions over the available patches.

We describe the range of approaches used to study both the ecological and evolutionary effects of interspecific competition, paying particular attention to experiments in the laboratory or field (e.g. substitutive, additive, response surface analysis) and natural experiments (e.g. comparing niche dimensions of species in sympatry and allopatry). The important question of whether a minimum amount of niche differentiation is required for stable coexistence is much easier to pose than answer.

The chapter concludes by acknowledging the need to consider not just the population dynamics of the competing populations but also the dynamics of the resources for which they are competing, if we wish to achieve a full understanding of interspecific competition and species coexistence.



Chapter 9

The Nature of Predation

9.1 Introduction: the types of predators

Consumers affect the distribution and abundance of the things they consume and vice versa, and these effects are of central importance in ecology. Yet, it is never an easy task to determine what the effects are, how they vary and why they vary. These topics will be dealt with in this and the next few chapters. We begin here by asking ‘What is the nature of predation?’, ‘What are the effects of predation on the predators themselves and on their prey?’ and ‘What determines where predators feed and what they feed on?’ In Chapter 10, we turn to the consequences of predation for the dynamics of predator and prey populations.

definition of predation

Predation, put simply, is consumption of one organism (the prey) by another organism (the predator), in which the prey is alive when the predator first attacks it. This excludes detritivory, the consumption of dead organic matter, which is discussed in its own right in Chapter 11. Nevertheless, it is a definition that encompasses a wide variety of interactions and a wide variety of ‘predators’.

taxonomic and functional classifications of predators

There are two main ways in which predators can be classified. Neither is perfect, but both can be useful. The most obvious classification is ‘taxonomic’: carnivores consume animals, herbivores consume plants and omnivores consume both (or, more correctly, prey from more than one trophic level – plants and herbivores, or herbivores and carnivores). An alternative, however, is a ‘functional’ classification of the type already outlined in Chapter 3. Here, there are four main types of predator: true predators, grazers, parasitoids and parasites (the last is divisible further into microparasites and macroparasites as explained in Chapter 12).

true predators

True predators kill their prey more or less immediately after attacking

them; during their lifetime they kill several or many different prey individuals, often consuming prey in their entirety. Most of the more obvious carnivores like tigers, eagles, coccinellid beetles and carnivorous plants are true predators, but so too are seed-eating rodents and ants, plankton-consuming whales, and so on.

Grazers also attack large numbers of prey during their lifetime, but they remove only part of each prey individual rather than the whole. Their effect on a prey individual, although typically harmful, is rarely lethal in the short term, and certainly never predictably lethal (in which case they would be true predators). Amongst the more obvious examples are the large vertebrate herbivores like sheep and cattle, but the flies that bite a succession of vertebrate prey, and leeches that suck their blood, are also undoubtedly grazers by this definition.

grazers

Parasites, like grazers, consume parts of their prey (their ‘host’), rather than the whole, and are typically harmful but rarely lethal in the short term. Unlike grazers, however, their attacks are concentrated on one or a very few individuals during their life. There is, therefore, an intimacy of association between parasites and their hosts that is not seen in true predators and grazers. Tapeworms, liver flukes, the measles virus, the tuberculosis bacterium and the flies and wasps that form mines and galls on plants are all obvious examples of parasites. There are also many plants, fungi and microorganisms that are parasitic on plants (often called ‘plant pathogens’), including the tobacco mosaic virus, the rusts and smuts and the mistletoes. Moreover, many herbivores may readily be thought of as parasites. For example, aphids extract sap from one or a very few individual plants with which they enter into intimate contact. Even caterpillars often rely on a single plant for their development. Plant pathogens, and animals parasitic on animals, will be dealt with together in Chapter 12. ‘Parasitic’ herbivores, like aphids and caterpillars, are dealt with here and in the next chapter, where we group them

parasites

together with true predators, grazers and parasitoids under the umbrella term ‘predator’.

parasitoids The parasitoids are a group of insects that belong mainly to the order Hymenoptera, but also include many Diptera. They are free-living as adults, but lay their eggs in, on or near other insects (or, more rarely, in spiders or woodlice). The larval parasitoid then develops inside or on its host. Initially, it does little apparent harm, but eventually it almost totally consumes the host and therefore kills it. An adult parasitoid emerges from what is apparently a developing host. Often, just one parasitoid develops from each host, but in some cases several or many individuals share a host. Thus, parasitoids are intimately associated with a single host individual (like parasites), they do not cause immediate death of the host (like parasites and grazers), but their eventual lethality is inevitable (like predators). For parasitoids, and also for the many herbivorous insects that feed as larvae on plants, the rate of ‘predation’ is determined very largely by the rate at which the adult females lay eggs. Each egg is an ‘attack’ on the prey or host, even though it is the larva that hatches from the egg that does the eating.

Parasitoids might seem to be an unusual group of limited general importance. However, it has been estimated that they account for 10% or more of the world’s species (Godfray, 1994). This is not surprising given that there are so many species of insects, that most of these are attacked by at least one parasitoid, and that parasitoids may in turn be attacked by parasitoids. A number of parasitoid species have been intensively studied by ecologists, and they have provided a wealth of information relevant to predation generally.

In the remainder of this chapter, we examine the nature of predation. We will look at the effects of predation on the prey individual (Section 9.2), the effects on the prey population as a whole (Section 9.3) and the effects on the predator itself (Section 9.4). In the cases of attacks by true predators and parasitoids, the effects on prey individuals are very straightforward: the prey is killed. Attention will therefore be placed in Section 9.2 on prey subject to grazing and parasitic attack, and herbivory will be the principal focus. Apart from being important in its own right, herbivory serves as a useful vehicle for discussing the subtleties and variations in the effects that predators can have on their prey.

Later in the chapter we turn our attention to the behavior of predators and discuss the factors that determine diet (Section 9.5) and where and when predators forage (Section 9.6). These topics are of particular interest in two broad contexts. First, foraging is an aspect of animal behavior that is subject to the scrutiny of evolutionary biologists, within the general field of ‘behavioral ecology’. The aim, put simply, is to try to understand how natural selection has favored particular patterns of behavior in particular circumstances (how, behaviorally, organisms match their environment). Second, the various aspects of predatory behavior can be seen as components that combine to influence the population

dynamics of both the predator itself and its prey. The population ecology of predation is dealt with much more fully in the next chapter.

9.2 Herbivory and individual plants: tolerance or defense

The effects of herbivory on a plant depend on which herbivores are involved, which plant parts are affected, and the timing of attack relative to the plant’s development. In some insect–plant interactions as much as 140 g, and in others as little as 3 g, of plant tissue are required to produce 1 g of insect tissue (Gavloski & Lamb, 2000a) – clearly some herbivores will have a greater impact than others. Moreover, leaf biting, sap sucking, mining, flower and fruit damage and root pruning are all likely to differ in the effect they have on the plant. Furthermore, the consequences of defoliating a germinating seedling are unlikely to be the same as those of defoliating a plant that is setting its own seed. Because the plant usually remains alive in the short term, the effects of herbivory are also crucially dependent on the response of the plant. Plants may show tolerance of herbivore damage or resistance to attack.

9.2.1 Tolerance and plant compensation

Plant compensation is a term that refers to the degree of tolerance exhibited by plants. If damaged plants have greater fitness than their undamaged

individual plants can compensate for herbivore effects

counterparts, they have *overcompensated*, and if they have lower fitness, they have *undercompensated* for herbivory (Strauss & Agrawal, 1999). Individual plants can compensate for the effects of herbivory in a variety of ways. In the first place, the removal of shaded leaves (with their normal rates of respiration but low rates of photosynthesis; see Chapter 3) may improve the balance between photosynthesis and respiration in the plant as a whole. Second, in the immediate aftermath of an attack from a herbivore, many plants compensate by utilizing reserves stored in a variety of tissues and organs or by altering the distribution of photosynthate within the plant. Herbivore damage may also lead to an increase in the rate of photosynthesis per unit area of surviving leaf. Often, there is compensatory regrowth of defoliated plants when buds that would otherwise remain dormant are stimulated to develop. There is also, commonly, a reduced death rate of surviving plant parts. Clearly, then, there are a number of ways in which individual plants compensate for the effects of herbivory (discussed further in Sections 9.2.3–9.2.5). But perfect compensation is rare. Plants are usually harmed by herbivores even though the compensatory reactions tend to counteract the harmful effects.

9.2.2 Defensive responses of plants

plants make defensive responses . . .

The evolutionary selection pressure exerted by herbivores has led to a variety of plant physical and chemical defenses that resist attack (see Sections 3.7.3 and 3.7.4). These may be present and effective continuously (constitutive defense) or increased production may be induced by attack (inducible defence) (Karban *et al.*, 1999). Thus, production of the defensive hydroxamic acid is induced when aphids (*Rhopalosiphum padi*) attack the wild wheat *Triticum uniaristatum* (Gianoli & Niemeier, 1997), and the prickles of dewberries on cattle-grazed plants are longer and sharper than those on ungrazed plants nearby (Abrahamson, 1975). Particular attention has been paid to rapidly inducible defenses, often the production of chemicals within the plant that inhibit the protease enzymes of the herbivores. These changes can occur within individual leaves, within branches or throughout whole tree canopies, and they may be detectable within a few hours, days or weeks, and last a few days, weeks or years; such responses have now been reported in more than 100 plant–herbivore systems (Karban & Baldwin, 1997).

. . . or do they?

There are, however, a number of problems in interpreting these responses (Schultz, 1988). First, are they ‘responses’ at all, or merely an incidental consequence of regrowth tissue having different properties from that removed by the herbivores? In fact, this issue is mainly one of semantics – if the metabolic responses of a plant to tissue removal happen to be defensive, then natural selection will favor them and reinforce their use. A further problem is much more substantial: are induced chemicals actually defensive in the sense of having an ecologically significant effect on the herbivores that seem to have induced them? Finally, and of most significance, are they truly defensive in the sense of having a measurable, positive impact on the plant making them, especially after the costs of mounting the response have been taken into account?

are herbivores really adversely affected? . . .

Fowler and Lawton (1985) addressed the second problem – ‘are the responses harmful to the herbivores?’ – by reviewing the effects of rapidly inducible plant defenses and found little clear-cut evidence that they are effective against insect herbivores, despite a widespread belief that they were. For example, they found that most laboratory studies revealed only small adverse effects (less than 11%) on such characters as larval development time and pupal weight, with many studies that claimed a larger effect being flawed statistically, and they argued that such effects may have negligible consequences for field populations. However, there are also a number of cases, many of which have been published since Fowler and Lawton’s review, in which the plant’s responses do seem to be genuinely harmful to the herbivores. When larch trees were defoliated by

the larch budmoth, *Zeiraphera diniana*, the survival and adult fecundity of the moths were reduced throughout the succeeding 4–5 years as a combined result of delayed leaf production, tougher leaves, higher fiber and resin concentration and lower nitrogen levels (Baltensweiler *et al.*, 1977). Another common response to leaf damage is early abscission (‘dropping off’) of mined leaves; in the case of the leaf-mining insect *Phyllonorycter* spp. on willow trees (*Salix lasiolepis*), early abscission of mined leaves was an important mortality factor for the moths – that is, the herbivores were harmed by the response (Preszler & Price, 1993). As a final example, a few weeks of grazing on the brown seaweed *Ascophyllum nodosum* by snails (*Littorina obtusata*) induces substantially increased concentrations of phlorotannins (Figure 9.1a), which reduce further snail grazing (Figure 9.1b). In this case, simple clipping of the plants did not have the same effect as the herbivore. Indeed, grazing by another herbivore, the isopod *Idotea granulosa*, also failed to induce the chemical defense. The snails can stay and feed on the same plant for long time periods (the isopods are much more mobile), so that induced responses that take time to develop can still be effective in reducing damage by snails.

The final question – ‘do plants benefit from their induced defensive responses?’ – has proved the most difficult to answer and only a few well

. . . and do plants really benefit?

designed field studies have been performed (Karban *et al.*, 1999). Agrawal (1998) estimated lifetime fitness of wild radish plants (*Raphanus sativus*) (as number of seeds produced multiplied by seed mass) assigned to one of three treatments: grazed plants (subject to grazing by the caterpillar of *Pieris rapae*), leaf damage controls (equivalent amount of biomass removed using scissors) and overall controls (undamaged). Damage-induced responses, both chemical and physical, included increased concentrations of defensive glucosinolates and increased densities of trichomes (hair-like structures). Earwigs (*Forficula* spp.) and other chewing herbivores caused 100% more leaf damage on the control and artificially leaf-clipped plants than on grazed plants and there were 30% more sucking green peach aphids (*Myzus persicae*) on the control and leaf-clipped plants (Figure 9.2a, b). Induction of resistance, caused by grazing by the *P. rapae* caterpillars, significantly increased the lifetime index of fitness by more than 60% compared to the control. However, leaf damage control plants (scissors) had 38% lower fitness than the overall controls, indicating the negative effect of tissue loss without the benefits of induction (Figure 9.2c).

This fitness benefit to wild radish occurred only in environments containing herbivores; in their absence, an induced defensive response was inappropriate and the plants suffered reduced fitness (Karban *et al.*, 1999). A similar fitness benefit has been shown in a field experiment involving wild tobacco (*Nicotiana attenuata*) (Baldwin, 1998). A specialist consumer of wild tobacco, the caterpillar of *Manduca sexta*, is remarkable in that it not only induces an accumulation of secondary metabolites and proteinase inhibitors when it feeds on wild tobacco, but it also induces the plants to

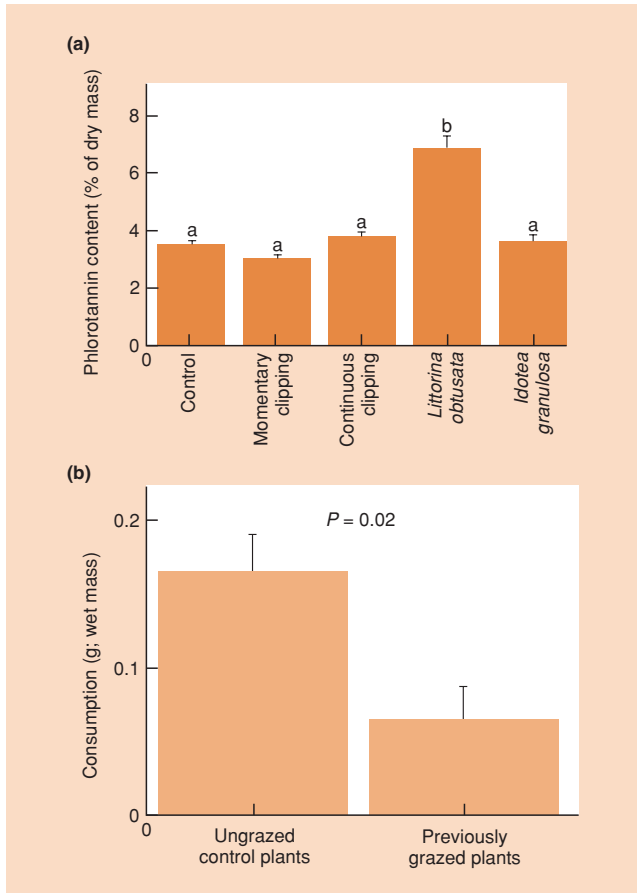


Figure 9.1 (a) Phlorotannin content of *Ascophyllum nodosum* plants after exposure to simulated herbivory (removing tissue with a hole punch) or grazing by real herbivores of two species. Means and standard errors are shown. Only the snail *Littorina obtusata* had the effect of inducing increased concentrations of the defensive chemical in the seaweed. Different letters indicate that means are significantly different ($P < 0.05$). (b) In a subsequent experiment, the snails were presented with algal shoots from the control and snail-grazed treatments in (a); the snails ate significantly less of plants with a high phlorotannin content. (After Pavia & Toth 2000.)

release volatile organic compounds that attract the generalist predatory bug *Geocoris pallens*, which feeds on the slow moving caterpillars (Kessler & Baldwin, 2004). Using molecular techniques, Zavala *et al.* (2004) were able to show that in the absence of herbivory, plant genotypes that produced little or no proteinase inhibitor grew faster and taller and produced more seed capsules than inhibitor-producing genotypes. Moreover, naturally occurring genotypes from Arizona that lacked the ability to produce proteinase inhibitors were damaged more, and sustained greater *Manduca* growth, in a laboratory experiment, compared with Utah inhibitor-producing genotypes (Glawe *et al.*, 2003).

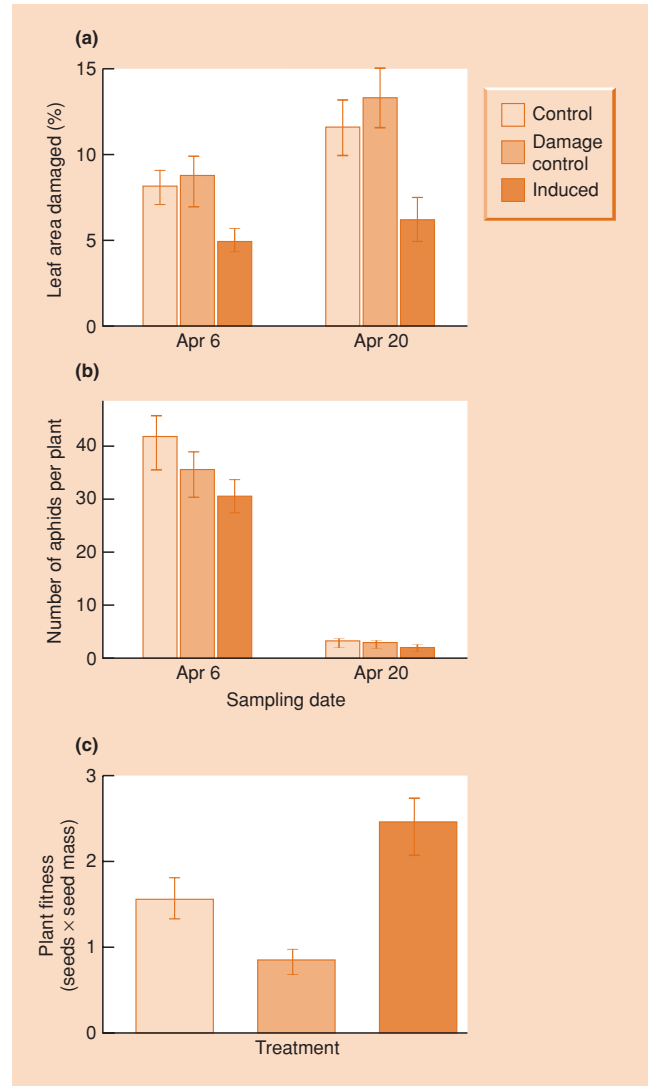


Figure 9.2 (a) Percentage of leaf area consumed by chewing herbivores and (b) number of aphids per plant, measured on two dates (April 6 and April 20) in three field treatments: overall control, damage control (tissue removed by scissors) and induced (caused by grazing of caterpillars of *Pieris rapae*). (c) The fitness of plants in the three treatments calculated by multiplying the number of seeds produced by the mean seed mass (in mg). (After Agrawal, 1998.)

It is clear from the wild radish and wild tobacco examples that the evolution of inducible (plastic) responses involves significant costs to the plant. We may expect inducible responses to be favored by selection only when past herbivory is a reliable predictor of future risk of herbivory *and* if the likelihood of herbivory is not constant (constant herbivory should select for a fixed defensive

phenotype that is best for that set of conditions) (Karban *et al.*, 1999). Of course, it is not only the costs of inducible defenses that can be set against fitness benefits. Constitutive defenses, such as spines, trichomes or defensive chemicals (particularly in the families Solanaceae and Brassicaceae), also have costs that have been measured (in phenotypes or genotypes lacking the defense) in terms of reductions in growth or the production of flowers, fruits or seeds (see review by Strauss *et al.*, 2002).

9.2.3 Herbivory, defoliation and plant growth

timing of herbivory is crucial

Despite a plethora of defensive structures and chemicals, herbivores still eat plants. Herbivory can stop plant growth, it can have a negligible effect on growth rate, and it can

do just about anything in between. Plant compensation may be a general response to herbivory or may be specific to particular herbivores. Gavloski and Lamb (2000b) tested these alternative hypotheses by measuring the biomass of two cruciferous plants *Brassica napus* and *Sinapis alba* in response to 0, 25 and 75% defoliation of seedling plants by three herbivore species with biting and chewing mouthparts – adult flea beetles *Phyllotreta cruciferae* and larvae of the moths *Plutella xylostella* and *Mamestra configurata*. Not surprisingly, both plant species compensated better for 25% than 75% defoliation. However, although defoliated to the same extent, both plants tended to compensate best for defoliation by the moth *M. configurata* and least for the beetle *P. cruciferae* (Figure 9.3). Herbivore-specific compensation may reflect plant responses to slightly different patterns of defoliation or different chemicals in saliva that suppress growth in contrasting ways (Gavloski & Lamb, 2000b).

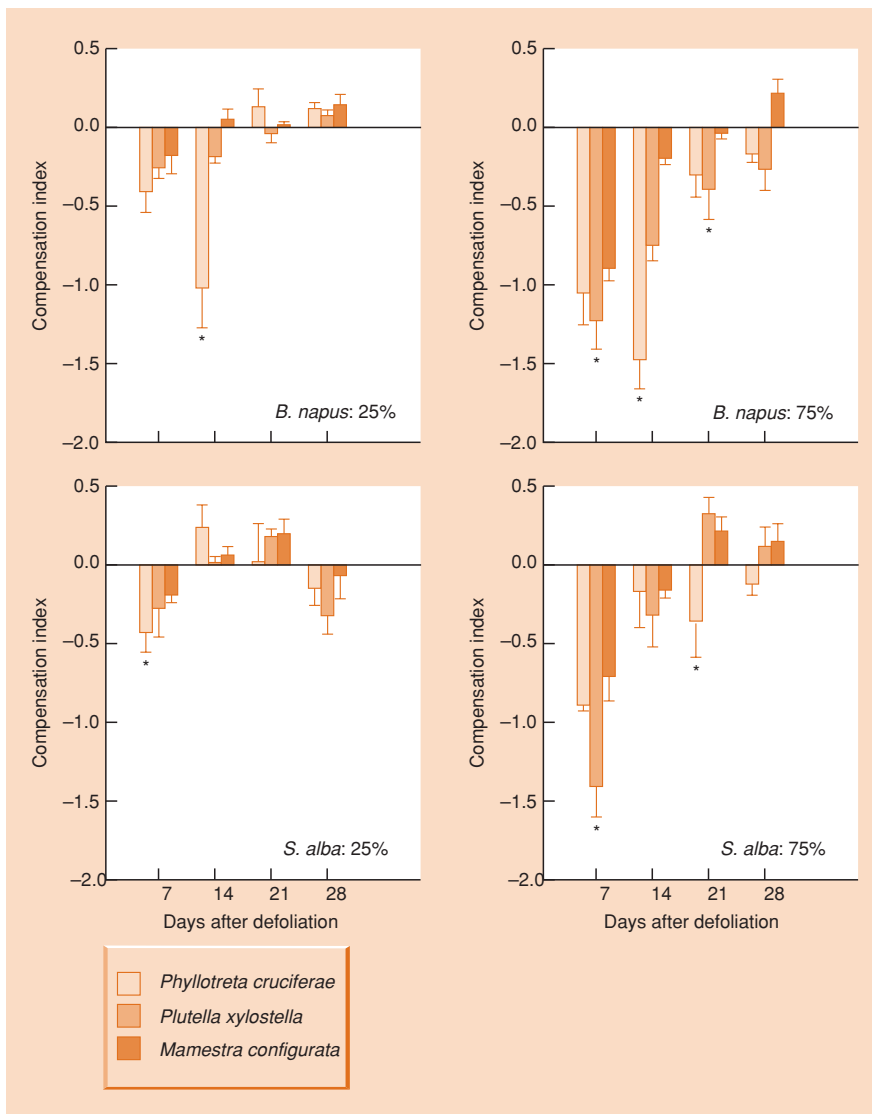


Figure 9.3 Compensation of leaf biomass (mean \pm SE: (\log_e biomass defoliated plant) – (\log_e of mean for control plants)) of *Brassica napus* and *Sinapis alba* seedlings with 25 or 75% defoliation by three species of insect (see key) in a controlled environment. On the vertical axis, zero equates to perfect compensation, negative values to undercompensation and positive values to overcompensation. Mean biomasses of defoliated plants that differ significantly from corresponding controls are indicated by an asterisk. (After Gavloski & Lamb, 2000b.)

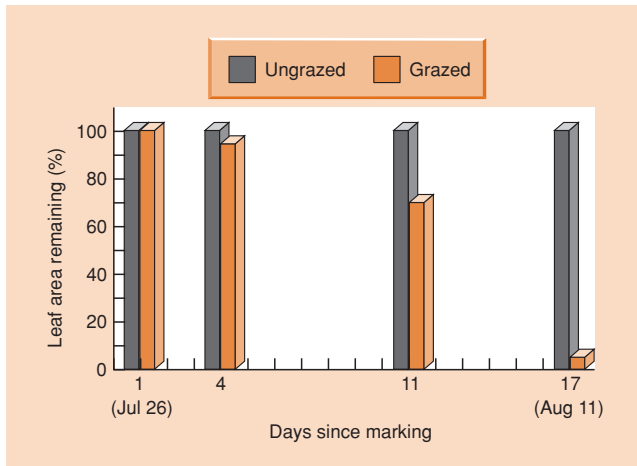


Figure 9.4 The survivorship of leaves on waterlily plants grazed by the waterlily leaf beetle was much lower than that on ungrazed plants. Effectively, all leaves had disappeared at the end of 17 days, despite the fact that ‘snapshot’ estimates of loss rates to grazing on grazed plants during this period suggested only around a 13% loss. (After Wallace & O’Hop, 1985.)

In the example above, compensation, which was generally complete by 21 days after defoliation, was associated with changes in root biomass consistent with the maintenance of a constant shoot : root ratio. Many plants compensate for herbivory in this way by altering the distribution of photosynthate in different parts of the plant. Thus, for example, Kosola *et al.* (2002) found that the concentration of soluble sugars in the young (white) fine roots of poplars (*Populus canadensis*) defoliated by gypsy moth caterpillars (*Lymantria dispar*) was much lower than in undefoliated trees. Older roots (>1 month in age), on the other hand, showed no significant effect of defoliation.

Often, there is considerable difficulty in assessing the real extent of defoliation, refoliation and hence net growth. Close monitoring of waterlily leaf beetles (*Pyrrhalta nymphaeae*) grazing on waterlilies (*Nuphar luteum*) revealed that leaves were rapidly removed, but that new leaves were also rapidly produced. More than 90% of marked leaves on grazed plants had disappeared within 17 days, while marked leaves on ungrazed plants were still completely intact (Figure 9.4). However, simple counts of leaves on grazed and ungrazed plants only indicated a 13% loss of leaves to the beetles, because of new leaf production on grazed plants.

grasses are particularly tolerant of grazing

The plants that seem most tolerant of grazing, especially vertebrate grazing, are the grasses. In most species, the meristem is almost at ground level amongst the basal leaf sheaths, and

this major point of growth (and regrowth) is therefore usually protected from grazers’ bites. Following defoliation, new leaves are produced using either stored carbohydrates or the photosynthate of surviving leaves, and new tillers are also often produced.

Grasses do not benefit directly from their grazers’ attentions. But it is likely that they are helped by grazers in their competitive interactions with other plants (which are more strongly affected by the grazers), accounting for the predominance of grasses in so many natural habitats that suffer intense vertebrate grazing. This is an example of the most widespread reason for herbivory having a more drastic effect on grazing-intolerant species than is initially apparent – the interaction between herbivory and plant competition (the range of possible consequences of which are discussed by Pacala & Crawley, 1992; see also Hendon & Briske, 2002). Note also that herbivores can have severe nonconsumptive effects on plants when they act as vectors for plant pathogens (bacteria, fungi and especially viruses) – what the herbivores take from the plant is far less important than what they give it! For instance, scolytid beetles feeding on the growing twigs of elm trees act as vectors for the fungus that causes Dutch elm disease. This killed vast numbers of elms in northeastern USA in the 1960s, and virtually eradicated them in southern England in the 1970s and early 1980s.

9.2.4 Herbivory and plant survival

Generally, it is more usual for herbivores to increase a plant’s susceptibility to mortality than to kill it outright. For example, although the flea beetle *Altica sublicata* reduced the growth rate of the sand-dune willow *Salix cordata* in both 1990 and 1991 (Figure 9.5), significant mortality as a result of drought stress only occurred in 1991. Then, however, susceptibility was strongly influenced by the herbivore: 80% of plants died in a high herbivory treatment (eight beetles per plant), 40% died at four beetles per plant, but none of the beetle-free control plants died (Bach, 1994).

mortality: the result of an interaction with another factor?

Repeated defoliation can have an especially drastic effect. Thus, a single defoliation of oak trees by the gypsy moth (*Lymantria dispar*) led to only a 5% mortality rate whereas three successive heavy defoliations led to mortality rates of up to 80% (Stephens, 1971). The mortality of established plants, however, is not necessarily associated with massive amounts of defoliation. One of the most extreme cases where the removal of a small amount of plant has a disproportionately profound effect is ring-barking of trees, for example by squirrels or porcupines. The cambial tissues and the phloem are torn away from the woody xylem, and the carbohydrate supply link between the leaves and the roots is broken. Thus, these pests of forestry plantations often kill young trees whilst removing very little tissue. Surface-feeding slugs can also do more damage to newly established grass populations than might be expected from the quantity of material they consume (Harper, 1977). The slugs chew through

repeated defoliation or ring-barking can kill

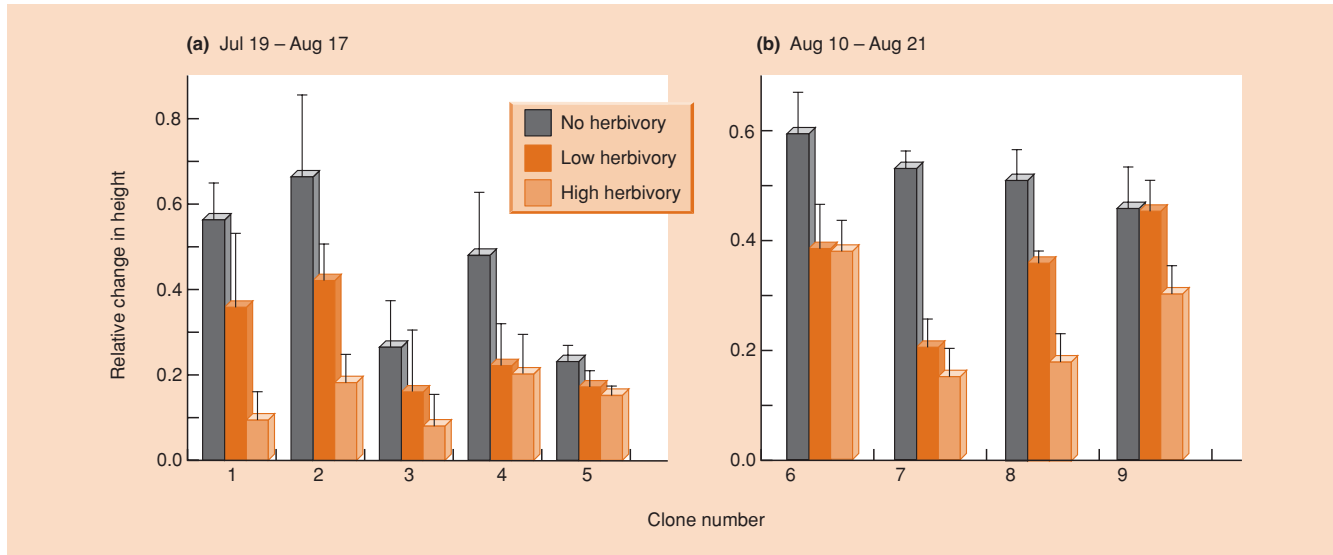


Figure 9.5 Relative growth rates (changes in height, with standard errors) of a number of different clones of the sand-dune willow, *Salix cordata*, (a) in 1990 and (b) in 1991, subjected either to no herbivory, low herbivory (four flea beetles per plant) or high herbivory (eight beetles per plant). (After Bach, 1994.)

the young shoots at ground level, leaving the felled leaves uneaten on the soil surface but consuming the meristematic region at the base of shoots from which regrowth would occur. They therefore effectively destroy the plant.

Predation of seeds, not surprisingly, has a predictably harmful effect on individual plants (i.e. the seeds themselves). Davidson *et al.* (1985) demonstrated dramatic impacts of seed-eating ants and rodents on the composition of seed banks of ‘annual’ plants in the deserts of southwestern USA and thus on the make up of the plant community.

9.2.5 Herbivory and plant fecundity

herbivores affect plant growth ...
... indirectly by reducing seed production ...

The effects of herbivory on plant fecundity are, to a considerable extent, reflections of the effects on plant growth: smaller plants bear fewer seeds. However, even when growth appears to be fully compensated, seed production

may nevertheless be reduced because of a shift of resources from reproductive output to shoots and roots. This was the case in the study shown in Figure 9.3 where compensation in growth was complete after 21 days but seed production was still significantly lower in the herbivore-damaged plants. Moreover, indirectly through its effect on leaf area, or by directly feeding on reproductive structures, herbivory can affect floral traits (corolla diameter, floral tube length, flower number) and have an adverse impact on pollination and seed set (Mothershead &

Marquis, 2000). Thus experimentally ‘grazed’ plants of *Oenothera macrocarpa* produced 30% fewer flowers and 33% fewer seeds.

Plants may also be affected more directly, by the removal or destruction of flowers, flower buds or seeds. Thus, caterpillars of the large blue butterfly *Maculinea rebeli* feed only in the flowers and on the fruits of the rare plant

... and directly by removing reproductive structures

Gentiana cruciata, and the number of seeds per fruit (70 compared to 120) is reduced where this specialist herbivore occurs (Kery *et al.*, 2001). Many studies, involving the artificial exclusion or removal of seed predators, have shown a strong influence of predispersal seed predation on recruitment and the density of attacked species. For example, seed predation was a significant factor in the pattern of increasing abundance of the shrub *Haplopappus squarrosus* along an elevational gradient from the Californian coast, where predispersal seed predation was higher, to the mountains (Louda, 1982); and restriction of the crucifer *Cardamine cordifolia* to shaded situations in the Rocky Mountains was largely due to much higher levels of predispersal seed predation in unshaded locations (Louda & Rodman, 1996).

It is important to realize, however, that many cases of ‘herbivory’ of reproductive tissues are actually mutualistic, benefitting both the herbivore and the plant (see Chapter 13). Animals that

much pollen and fruit herbivory benefits the plant

‘consume’ pollen and nectar usually transfer pollen inadvertently from plant to plant in the process; and there are many fruit-eating animals that also confer a net benefit on both the parent

plant and the individual seed within the fruit. Most vertebrate fruit-eaters, in particular, either eat the fruit but discard the seed, or eat the fruit but expel the seed in the feces. This disperses the seed, rarely harms it and frequently enhances its ability to germinate.

Insects that attack fruit or developing fruit, on the other hand, are very unlikely to have a beneficial effect on the plant. They do nothing to enhance dispersal, and they may even make the fruit less palatable to vertebrates. However, some large animals that normally kill seeds can also play a part in dispersing them, and they may therefore have at least a partially beneficial effect. There are some 'scatter-hoarding' species, like certain squirrels, that take nuts and bury them at scattered locations; and there are other 'seed-caching' species, like some mice and voles, that collect scattered seeds into a number of hidden caches. In both cases, although many seeds are eaten, the seeds are dispersed, they are hidden from other seed predators and a number are never relocated by the hoarder or cacher (Crawley, 1983).

Herbivores also influence fecundity in a number of other ways. One of the most common responses to herbivore attack is a delay in flowering. For instance, in longer lived semelparous species, herbivory frequently delays flowering for 1 year or more, and this typically increases the longevity of such plants since

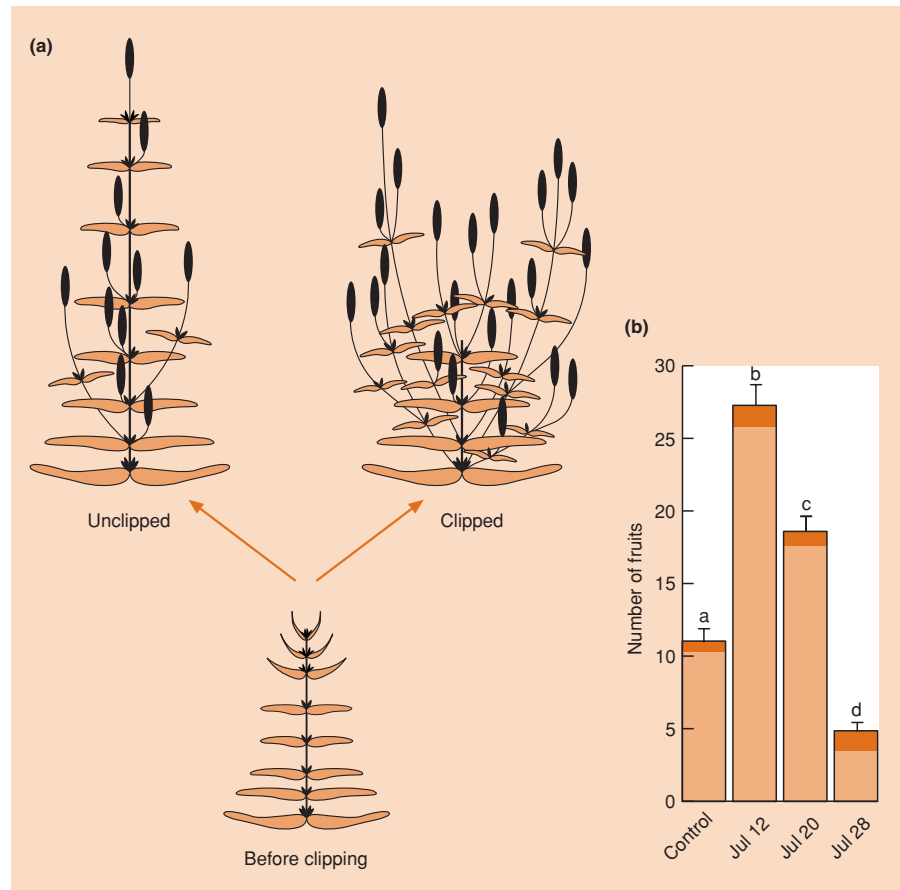
death almost invariably follows their single burst of reproduction (see Chapter 4). *Poa annua* on a lawn can be made almost immortal by mowing it at weekly intervals, whereas in natural habitats, where it is allowed to flower, it is commonly an annual – as its name implies.

Generally, the timing of defoliation is critical in determining the effect on plant fecundity. If leaves are removed before inflorescences are formed, then the extent to which fecundity is depressed clearly depends on the extent to which the plant is able to compensate. Early defoliation of a plant with sequential leaf production may have a negligible effect on fecundity; but where defoliation takes place later, or where leaf production is synchronous, flowering may be reduced or even inhibited completely. If leaves are removed after the inflorescence has been formed, the effect is usually to increase seed abortion or to reduce the size of individual seeds.

An example where timing is important is provided by field gentians (*Gentianella campestris*). When herbivory on this biennial plant is simulated by clipping to remove half its biomass (Figure 9.6a), the outcome depends on the timing of the clipping (Figure 9.6b). Fruit production was much increased over controls if clipping

the timing of herbivory is critical

Figure 9.6 (a) Clipping of field gentians to simulate herbivory causes changes in the architecture and numbers of flowers produced. (b) Production of mature (open histograms) and immature fruits (black histograms) of unclipped control plants and plants clipped on different occasions from July 12 to 28, 1992. Means and standard errors are shown and all means are significantly different from each other ($P < 0.05$). Plants clipped on July 12 and 20 developed significantly more fruits than unclipped controls. Plants clipped on July 28 developed significantly fewer fruits than controls. (After Lennartsson *et al.*, 1998).



occurred between 1 and 20 July, but if clipping occurred later than this, fruit production was less in the clipped plants than in the unclipped controls. The period when the plants show compensation coincides with the time when damage by herbivores normally occurs.

9.2.6 A postscript: antipredator chemical defenses in animals

animals also defend themselves

It should not be imagined that antipredator chemical defenses are restricted to plants. A variety of constitutive animal chemical defenses were described in Chapter 3 (see Section 3.7.4), including plant defensive chemicals sequestered by herbivores from their food plants (see Section 3.7.4). Chemical defenses may be particularly important in modular animals, such as sponges, which lack the ability to escape from their predators. Despite their high nutritional value and lack of physical defenses, most marine sponges appear to be little affected by predators (Kubaneck *et al.*, 2002). In recent years, several triterpene glycosides have been extracted from sponges, including from *Ectyoplasia ferox* in the Caribbean. In a field study, crude extracts of refined triterpene glycosides from this sponge were presented in artificial food substrates to natural assemblages of reef fishes in the Bahamas. Strong antipredatory affects were detected when compared to control substrates (Figure 9.7). It is of interest that the triterpene glycosides also adversely affected competitors of the sponge, including 'fouling' organisms that overgrow them (bacteria, invertebrates and algae) and other sponges (an example of allelopathy – see Section 8.3.2). All these enemies were apparently deterred by surface contact with the chemicals rather than by water-borne effects (Kubaneck *et al.*, 2002).

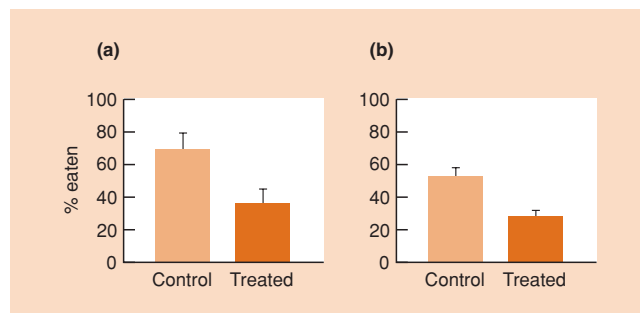


Figure 9.7 Results of field assays assessing antipredatory effects of compounds from the sponge *Ectyoplasia ferox* against natural assemblages of reef fish in the Bahamas. Means (+ SE) are shown for percentages of artificial food substrates eaten in controls (containing no sponge extracts) in comparison with: (a) substrates containing a crude sponge extract (t -test, $P = 0.036$) and (b) substrates containing triterpene glycosides from the sponge ($P = 0.011$). (After Kubaneck *et al.*, 2002.)

9.3 The effect of predation on prey populations

Returning now to predators in general, it may seem that since the effects of predators are harmful to individual prey, the immediate effect of predation on a population of prey must also be predictably harmful. However, these effects are not always so predictable, for one or both of two important reasons. In the first place, the individuals that are killed (or harmed) are not always a random sample of the population as a whole, and may be those with the lowest potential to contribute to the population's future. Second, there may be compensatory changes in the growth, survival or reproduction of the surviving prey: they may experience reduced competition for a limiting resource, or produce more offspring, or other predators may take fewer of the prey. In other words, whilst predation is bad for the prey that get eaten, it may be good for those that do not. Moreover, predation is least likely to affect prey dynamics if it occurs at a stage of the prey's life cycle that does not have a significant effect, ultimately, on prey abundance.

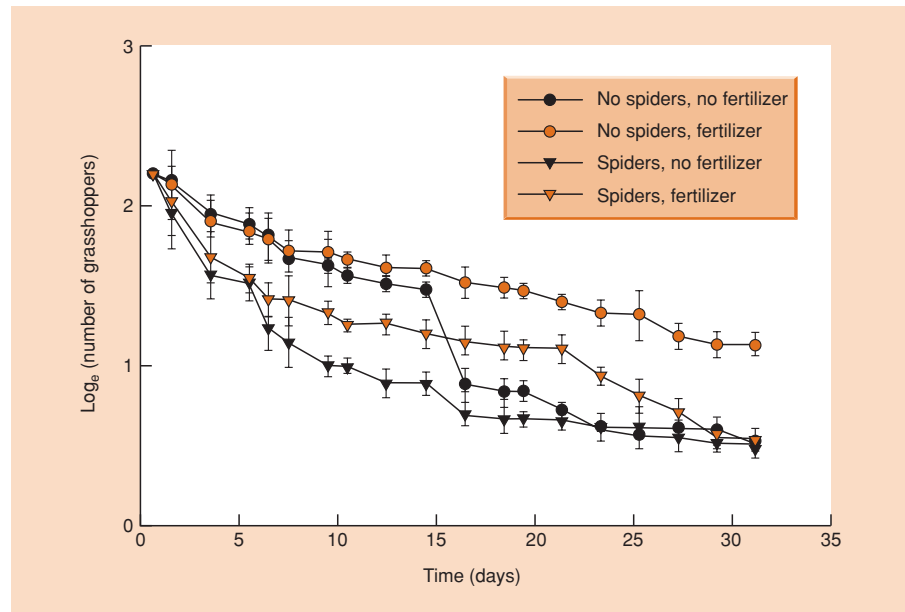
To deal with the second point first, if, for example, plant recruitment is not limited by the number of seeds produced, then insects that reduce seed production are unlikely to have an important effect on plant abundance (Crawley, 1989). For instance, the weevil *Rhinocyllus conicus* does not reduce recruitment of the nodding thistle, *Carduus nutans*, in southern France despite inflicting seed losses of over 90%. Indeed, sowing 1000 thistle seeds per square meter also led to no observable increase in the number of thistle rosettes. Hence, recruitment appears not to be limited by the number of seeds produced; although whether it is limited by subsequent predation of seeds or early seedlings, or the availability of germination sites, is not clear (Crawley, 1989). (However, we have seen in other situations (see Section 9.2.5) that predispersal seed predation can profoundly affect seedling recruitment, local population dynamics and variation in relative abundance along environmental gradients and across microhabitats.)

The impact of predation is often limited by compensatory reactions amongst the survivors as a result of reduced intraspecific competition. Thus, in a classic experiment in which large numbers of woodpigeons (*Columba palumbus*) were shot, the overall level of winter mortality was not increased, and stopping the shooting led to no increase in pigeon abundance (Murton *et al.*, 1974). This was because the number of surviving pigeons was determined ultimately not by shooting but by food availability, and so when shooting reduced density, there were compensatory reductions in intraspecific competition and in natural mortality, as well as density-dependent immigration of birds moving in to take advantage of unexploited food.

predation may occur at a demographically unimportant stage

compensatory reactions amongst survivors

Figure 9.8 Trajectories of numbers of grasshoppers surviving (mean \pm SE) for fertilizer and predation treatment combinations in a field experiment involving caged plots in the Arapaho Prairie, Nebraska, USA. (After Oedekoven & Joern, 2000.)



effects ameliorated by reduced competition

Indeed, whenever density is high enough for intraspecific competition to occur, the effects of predation on a population should be ameliorated by the consequent reductions in intraspecific competition. Outcomes of predation may, therefore, vary with relative food availability. Where food quantity or quality is higher, a given level of predation may not lead to a compensatory response because prey are not food-limited. This hypothesis was tested by Oedekoven and Joern (2000) who monitored grasshopper (*Ageneotettix deorum*) survivorship in caged prairie plots subject to fertilization (or not) to increase food quality in the presence or absence of lycosid spiders (*Schizocoza* spp.). With ambient food quality (no fertilizer, black symbols), spider predation and food limitation were compensatory: the same numbers of grasshoppers were recovered at the end of the 31-day experiment (Figure 9.8). However, with higher food quality (nitrogen fertilizer added, colored symbols), spider predation reduced the numbers surviving compared to the no-spider control: a noncompensatory response. Under ambient conditions after spider predation, the surviving grasshoppers encountered more food per capita and lived longer as a result of reduced competition. However, grasshoppers were less food-limited when food quality was higher so that after predation the release of additional per capita food did not promote survivorship (Oedekoven & Joern, 2000).

predatory attacks are often directed at the weakest prey

Turning to the nonrandom distribution of predators' attention within a population of prey, it is likely, for example, that predation by many large carnivores is focused on the old (and infirm), the young (and naive) or the sick. For instance, a study

in the Serengeti found that cheetahs and wild dogs killed a disproportionate number from the younger age classes of Thomson's gazelles (Figure 9.9a), because: (i) these young animals were easier to catch (Figure 9.9b); (ii) they had lower stamina and running speeds; (iii) they were less good at outmaneuvering the predators (Figure 9.9c); and (iv) they may even have failed to recognize the predators (FitzGibbon & Fanshawe, 1989; FitzGibbon, 1990). Yet these young gazelles will also have been making no reproductive contribution to the population, and the effects of this level of predation on the prey population will therefore have been less than would otherwise have been the case.

Similar patterns may also be found in plant populations. The mortality of mature *Eucalyptus* trees in Australia, resulting from defoliation by the sawfly *Paropsis atomaria*, was restricted almost entirely to weakened trees on poor sites, or to trees that had suffered from root damage or from altered drainage following cultivation (Carne, 1969).

Taken overall, then, it is clear that the step from noting that individual prey are harmed by individual predators to demonstrating that prey abundance

difficulties of demonstrating effects on prey populations

is adversely affected is not an easy one to take. Of 28 studies in which herbivorous insects were experimentally excluded from plant communities using insecticides, 50% provided evidence of an effect on plants at the population level (Crawley, 1989). As Crawley noted, however, such proportions need to be treated cautiously. There is an almost inevitable tendency for 'negative' results (no population effect) to go unreported, on the grounds of there being 'nothing' to report. Moreover, the exclusion studies often took 7 years or more to show any impact on the plants: it may be that many of the 'negative' studies were simply given up too early.

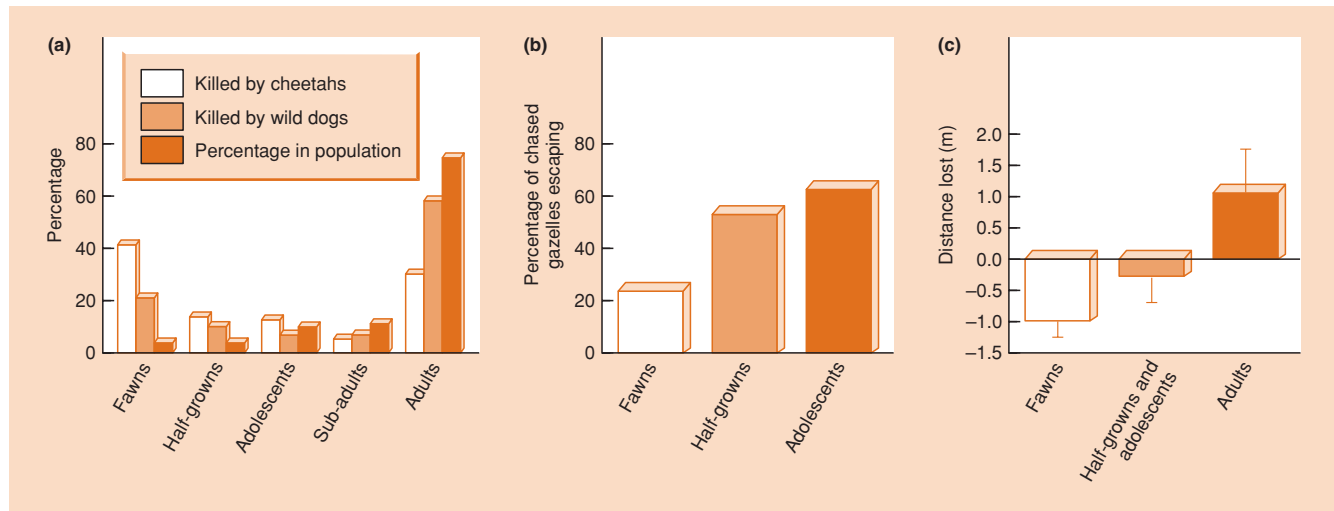


Figure 9.9 (a) The proportions of different age classes (determined by tooth wear) of Thomson's gazelles in cheetah and wild dog kills is quite different from their proportions in the population as a whole. (b) Age influences the probability for Thomson's gazelles of escaping when chased by cheetahs. (c) When prey (Thomson's gazelles) 'zigzag' to escape chasing cheetahs, prey age influences the mean distance lost by the cheetahs. (After FitzGibbon & Fanshawe, 1989; FitzGibbon, 1990.)

Many more recent investigations have shown clear effects of seed predation on plant abundance (e.g. Kelly & Dyer, 2002; Maron *et al.*, 2002).

9.4 Effects of consumption on consumers

consumers often need to exceed a threshold of consumption

The beneficial effects that food has on individual predators are not difficult to imagine. Generally speaking, an increase in the amount of food consumed leads to increased rates of growth, development and birth, and decreased rates of mortality. This, after all, is implicit in any discussion of intraspecific competition amongst consumers (see Chapter 5): high densities, implying small amounts of food per individual, lead to low growth rates, high death rates, and so on. Similarly, many of the effects of migration previously considered (see Chapter 6) reflect the responses of individual consumers to the distribution of food availability. However, there are a number of ways in which the relationships between consumption rate and consumer benefit can be more complicated than they initially appear. In the first place, all animals require a certain amount of food simply for maintenance and unless this threshold is exceeded the animal will be unable to grow or reproduce, and will therefore be unable to contribute to future generations. In other words, low consumption rates, rather than leading to a small benefit to the consumer, simply alter the rate at which the consumer starves to death.

At the other extreme, the birth, growth and survival rates of individual consumers cannot be expected to rise indefinitely as food availability is increased. Rather, the consumers become satiated. Consumption rate eventually reaches a plateau, where it becomes independent of the amount of food available, and benefit to consumers therefore also reaches a plateau. Thus, there is a limit to the amount that a particular consumer population can eat, a limit to the amount of harm that it can do to its prey population at that time, and a limit to the extent by which the consumer population can increase in size. This is discussed more fully in Section 10.4.

consumers may become satiated

The most striking example of whole populations of consumers being satiated simultaneously is provided by the many plant species that have mast years. These are occasional years in which there is synchronous production of a large volume of seed, often across a large geographic area, with a dearth of seeds produced in the years in between (Herrera *et al.*, 1998; Koenig & Knops, 1998; Kelly *et al.*, 2000). This is seen particularly often in tree species that suffer generally high intensities of seed predation (Silvertown, 1980) and it is therefore especially significant that the chances of escaping seed predation are likely to be much higher in mast years than in other years. Masting seems to be especially common in the New Zealand flora (Kelly, 1994) where it has also been reported for tussock grass species (Figure 9.10). The individual predators of seeds are satiated in mast years, and the populations of predators cannot increase in abundance rapidly enough to exploit the glut. This

most years and the satiation of seed predators

Figure 9.10 The flowering rate for five species of tussock grass (*Chionochloa*) between 1973 and 1996 in Fiordland National Park, New Zealand. Mast years are highly synchronized in the five species, seemingly in response to high temperatures in the previous season, when flowering is induced. (After McKone *et al.*, 1998.)

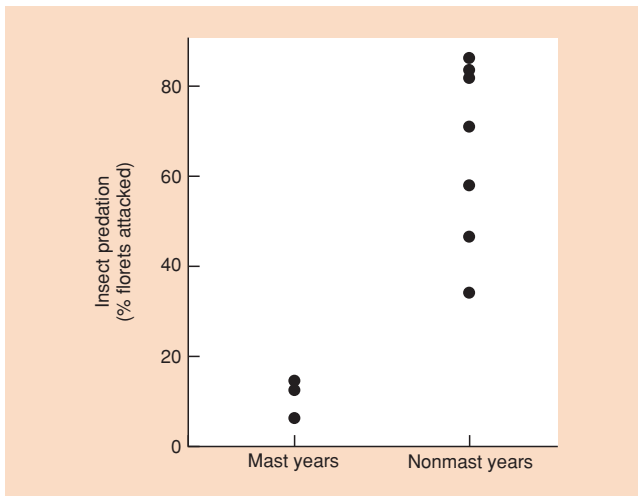
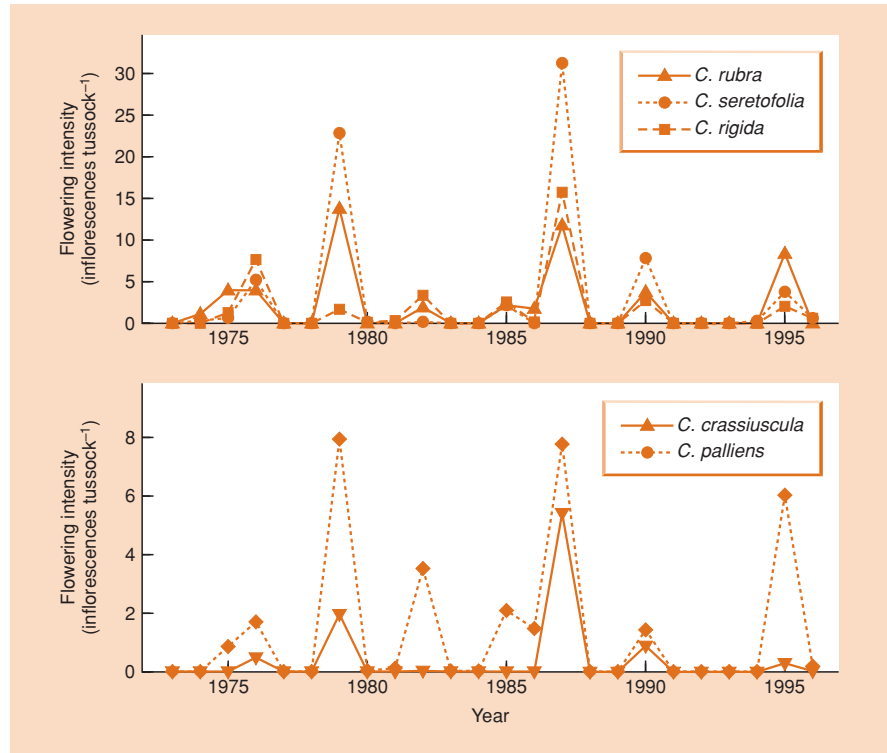


Figure 9.11 Insect predation on florets of *Chionochloa palliens* in mast ($n = 3$) and nonmast years ($n = 7$) from 1988 to 1997 at Mount Hutt, New Zealand. A mast year is defined here as one with greater than 10 times as many florets produced per tussock than in the previous year. The significant difference in insect damage supports the hypothesis that the function of masting is to satiate seed predators. (After McKone *et al.*, 1998.)

is illustrated in Figure 9.11 where the percentage of florets of the grass *Chionochloa pallens* attacked by insects remains below 20% in mast years but ranges up to 80% or more in nonmast years. The fact that *C. pallens* and four other species of *Chionochloa* show strong synchrony in masting is likely to result in an increased benefit to each species in terms of escaping seed predation in mast years.

On the other hand, the production of a mast crop makes great demands on the internal resources of a plant. A spruce tree in a mast year averages 38% less annual growth than in other years, and the annual ring increment in forest trees may be reduced by as much during a mast year as by a heavy attack of defoliating caterpillars. The years of seed famine are therefore essentially years of plant recovery.

As well as illustrating the potential importance of predator satiation, the example of masting highlights a further point relating to timescales. The seed predators are unable to extract the maximum benefit from (or do the maximum harm to) the mast crop because their generation times are too long. A hypothetical seed predator population that could pass through several generations during a season would be able to increase exponentially and explosively on the mast crop and destroy it. Generally speaking, consumers with relatively short generation times tend to closely track fluctuations in the quantity or abundance of their food or

a consumer's numerical response is limited by its generation time . . .

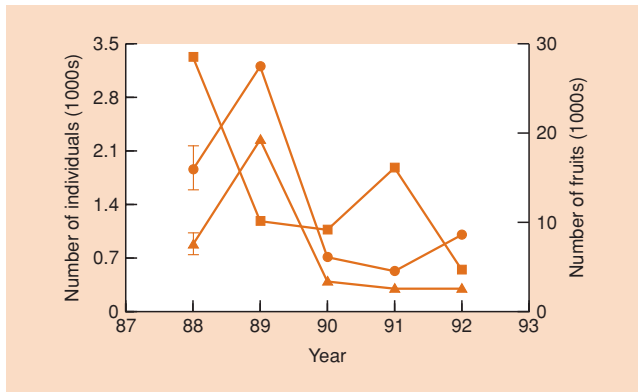


Figure 9.12 Fluctuations in the fruit production of *Asphodelus* (■) and the number of *Capsodes* nymphs (●) and adults (▲) at a study site in the Negev desert, Israel. (After Ayal, 1994.)

prey, whereas consumers with relatively long generation times take longer to respond to increases in prey abundance, and longer to recover when reduced to low densities.

... as illustrated by
desert interactions

The same phenomenon occurs in desert communities, where year-to-year variations in precipitation can be both considerable and unpredictable, leading to similar year-to-year variation in the productivity of many desert plants. In the rare years of high productivity, herbivores are typically at low abundance following one or more years of low plant productivity. Thus, the herbivores are likely to be satiated in such years, allowing plant populations to add considerably to their reserves, perhaps by augmenting their buried seed banks or their underground storage organs (Ayal, 1994). The example of fruit production by *Asphodelus ramosus* in the Negev desert in Israel is shown in Figure 9.12. The mirid bug, *Capsodes infuscatus*, feeds on *Asphodelus*, exhibiting a particular preference for the developing flowers and young fruits. Potentially, therefore, it can have a profoundly harmful effect on the plant's fruit production. But it only passes through one generation per year. Hence, its abundance tends never to match that of its host plant (Figure 9.12). In 1988 and 1991, fruit production was high but mirid abundance was relatively low: the reproductive output of the mirids was therefore high (3.7 and 3.5 nymphs per adult, respectively), but the proportion of fruits damaged was relatively low (0.78 and 0.66). In 1989 and 1992, on the other hand, when fruit production had dropped to much lower levels, the proportion of fruits damaged was much higher (0.98 and 0.87) and the reproductive output was lower (0.30 nymphs per adult in 1989; unknown in 1992). This suggests that herbivorous insects, at least, may have a limited ability to affect plant population dynamics in desert communities, but that the potential is much greater for the dynamics of herbivorous insects to be affected by their food plants (Ayal, 1994).

Chapter 3 stressed that the quantity of food consumed may be less important than its quality. In fact, food quality, which has both positive aspects (like the concentrations of nutrients) and negative aspects (like the concentrations of toxins), can only sensibly be defined in terms of the effects of the food on the animal that eats it; and this is particularly pertinent in the case of herbivores. For instance, we saw in Figure 9.8 how even in the presence of predatory spiders, enhanced food quality led to increased survivorship of grasshoppers. Along similar lines, Sinclair (1975) examined the effects of grass quality (protein content) on the survival of wildebeest in the Serengeti of Tanzania. Despite selecting protein-rich plant material (Figure 9.13a), the wildebeest consumed food in the dry season that contained well below the level of protein necessary even for maintenance (5–6% of crude protein); and to judge by the depleted fat reserves of dead males (Figure 9.13b), this was an important cause of mortality. Moreover, it is highly relevant that the protein requirements of females during late pregnancy and lactation (December–May in the wildebeest) are three to four times higher than the normal. It is therefore clear that the shortage of high-quality food (and not just food shortage *per se*) can have a drastic effect on the growth, survival and fecundity of a consumer. In the case of herbivores especially, it is possible for an animal to be apparently surrounded by its food whilst still experiencing a food shortage. We can see the problem if we imagine that we ourselves are provided with a perfectly balanced diet – diluted in an enormous swimming pool. The pool contains everything we need, and we can see it there before us, but we may very well starve to death before we can drink enough water to extract enough nutrients to sustain ourselves. In a similar fashion, herbivores may frequently be confronted with a pool of available nitrogen that is so dilute that they have difficulty processing enough material to extract what they need. Outbreaks of herbivorous insects may then be associated with rare elevations in the concentration of available nitrogen in their food plants (see Section 3.7.1), perhaps associated with unusually dry or, conversely, unusually waterlogged conditions (White, 1993). Consumers obviously need to acquire resources – but, to benefit from them fully they need to acquire them in appropriate quantities and in an appropriate form. The behavioral strategies that have evolved in the face of the pressures to do this are the main topic of the next two sections.

food quality rather than quantity can be of paramount importance

9.5 Widths and compositions of diets

Consumers can be classified as either monophagous (feeding on a single prey type), oligophagous (few prey types) or polyphagous (many prey types). An equally useful distinction is

range and classification of diet widths

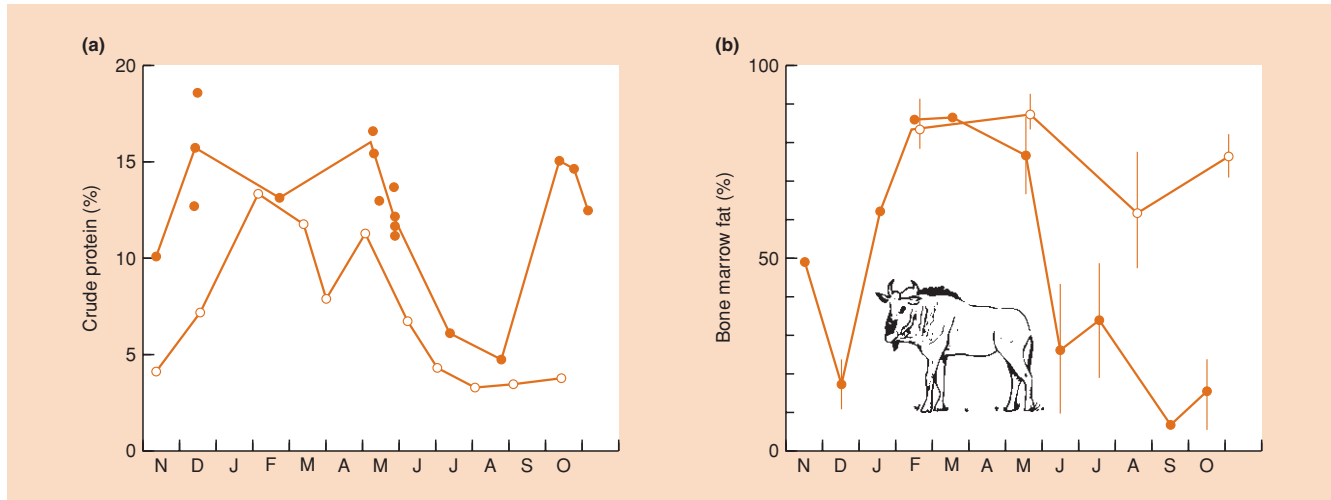


Figure 9.13 (a) The quality of food measured as percentage crude protein available to (○) and eaten by (●) wildebeest in the Serengeti during 1971. Despite selection ('eaten' > 'available'), the quality of food eaten fell during the dry season below the level necessary for the maintenance of nitrogen balance (5–6% of crude protein). (b) The fat content of the bone marrow of the live male population (○) and those found dead from natural causes (●). Vertical lines, where present, show 95% confidence limits. (After Sinclair, 1975.)

between specialists (broadly, monophages and oligophages) and generalists (polyphages). Herbivores, parasitoids and true predators can all provide examples of monophagous, oligophagous and polyphagous species. But the distribution of diet widths differs amongst the various types of consumer. True predators with specialized diets do exist (for instance the snail kite *Rostrhamus sociabilis* feeds almost entirely on snails of the genus *Pomacea*), but most true predators have relatively broad diets. Parasitoids, on the other hand, are typically specialized and may even be monophagous. Herbivores are well represented in all categories, but whilst grazing and 'predatory' herbivores typically have broad diets, 'parasitic' herbivores are very often highly specialized. For instance, Janzen (1980) examined 110 species of beetle that feed as larvae inside the seeds of dicotyledonous plants in Costa Rica ('parasitizing' them) and found that 83 attacked only one plant species, 14 attacked only two, nine attacked three, two attacked four, one attacked six and one attacked eight of the 975 plants in the area.

9.5.1 Food preferences

preference is defined by comparing diet with 'availability'

It must not be imagined that polyphagous and oligophagous species are indiscriminate in what they choose from their acceptable range. On the contrary, some degree of preference is almost always apparent. An animal is said to exhibit a preference for a particular type of food when the proportion of that type in the animal's diet is higher than its proportion in the animal's environment. To measure

food preference in nature, therefore, it is necessary not only to examine the animal's diet (usually by the analysis of gut contents) but also to assess the 'availability' of different food types. Ideally, this should be done not through the eyes of the observer (i.e. not by simply sampling the environment), but through the eyes of the animal itself.

A food preference can be expressed in two rather different contexts. There can be a preference for items that are the most valuable amongst those available *or* for items that provide an integral part of a mixed and balanced diet. These will be referred to as ranked and balanced preferences, respectively. In the terms of Chapter 3 (Section 3.8), where resources were classified, individuals exhibit ranked preferences in discriminating between resource types that are 'perfectly substitutable' and exhibit balanced preferences between resource types that are 'complementary'.

Ranked preferences are usually seen most clearly amongst carnivores. For instance, Figure 9.14 shows two examples in which carnivores actively selected prey items that were the most profitable in terms of energy intake per unit time spent dealing with (or 'handling') prey. Results such as these reflect the fact that a carnivore's food often varies little in composition (see Section 3.7.1), but may vary in size or accessibility. This allows a single measure (like 'energy gained per unit handling time') to be used to characterize food items, and it therefore allows food items to be ranked. In other words, Figure 9.14 shows consumers exhibiting an active preference for food of a high rank.

ranked preferences predominate when food items can be classified on a single scale ...

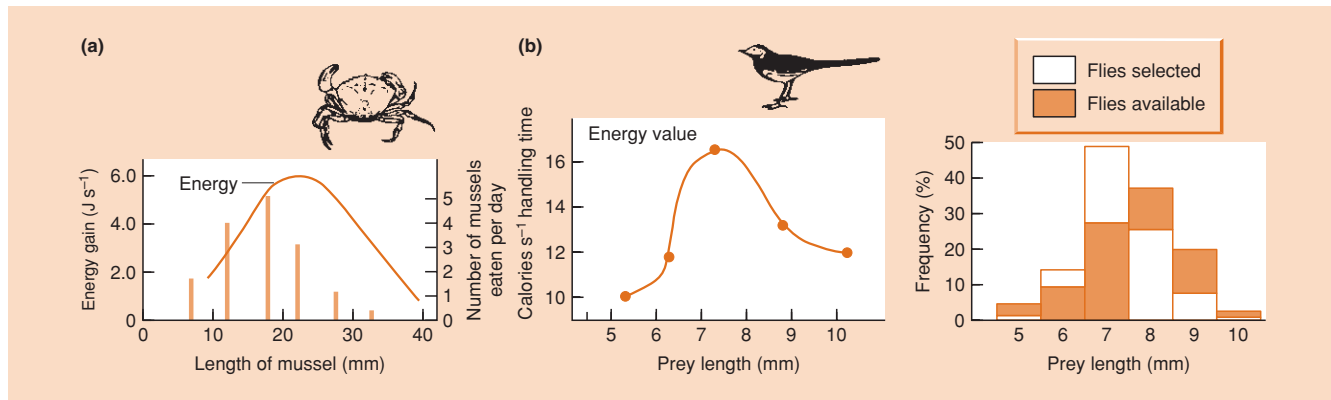


Figure 9.14 Predators eating ‘profitable’ prey, i.e. predators showing a preponderance in their diet for those prey items that provide them with the most energy. (a) When crabs (*Carcinus maenas*) were presented with equal quantities of six size classes of mussels (*Mytilus edulis*), they tended to show a preference for those providing the greatest energy gain (energy per unit handling time). (After Elner & Hughes, 1978.) (b) Pied wagtails (*Motacilla alba yarrellii*) tended to select, from scatophagid flies available, those providing the greatest energy gain per unit handling time. (After Davies, 1977; Krebs, 1978.)

... but many consumers show a combination of ranked and balanced preferences

For many consumers, however, especially herbivores and omnivores, no simple ranking is appropriate, since none of the available food items matches the nutritional requirements of the consumer. These requirements can therefore only be satisfied either by eating large quantities of food, and eliminating much of it in order to get a sufficient quantity of the nutrient in most limited supply (for example aphids and scale insects excrete vast amounts of carbon in honeydew to get sufficient nitrogen from plant sap), or by eating a combination of food items that between them match the consumer’s requirements. In fact, many animals exhibit both sorts of response. They select food that is of generally high quality (so the proportion eliminated is minimized), but they also select items to meet specific requirements. For instance, sheep and cattle show a preference for high-quality food, selecting leaves in preference to stems, green matter in preference to dry or old material, and generally selecting material that is higher in nitrogen, phosphorus, sugars and gross energy, and lower in fiber, than what is generally available. In fact, all generalist herbivores appear to show rankings in the rate at which they eat different food plants when given a free choice in experimental tests (Crawley, 1983).

mixed diets can be favored for a variety of reasons

On the other hand, a balanced preference is also quite common. For instance, the plate limpet, *Acmaea scutum*, selects a diet of two species of encrusting microalgae that contains 60% of one species and 40% of the other, almost irrespective of the proportions in which they are available (Kitting, 1980). Whilst caribou, which survive on lichen through the winter, develop a

sodium deficiency by the spring that they overcome by drinking seawater, eating urine-contaminated snow and gnawing shed antlers (Staaland *et al.*, 1980). We have only to look at ourselves to see an example in which ‘performance’ is far better on a mixed diet than on a pure diet of even the ‘best’ food.

There are two other important reasons why a mixed diet may be favored. First, consumers may accept low-quality items simply because, having encountered them, they have more to gain by eating them (poor as they are) than by ignoring them and continuing to search. This is discussed in detail in Section 9.5.3. Second, consumers may benefit from a mixed diet because each food type may contain a different undesirable toxic chemical. A mixed diet would then keep the concentrations of all of these chemicals within acceptable limits. It is certainly the case that toxins can play an important role in food preference. For instance, dry matter intake by Australian ringtail possums (*Pseudocheirus peregrinus*) feeding on *Eucalyptus* tree leaves was strongly negatively correlated with the concentration of sideroxylyonal, a toxin found in *Eucalyptus* leaves, but was not related to nutritional characteristics such as nitrogen or cellulose (Lawler *et al.*, 2000).

Overall, however, it would be quite wrong to give the impression that all preferences have been clearly linked with one explanation or another. For example, Thompson (1988) reviewed the relationship between the oviposition preferences of phytophagous insects and the performance of their offspring on the selected food plants in terms of growth, survival and reproduction. A number of studies have shown a good association (i.e. females preferentially oviposit on plants where their offspring perform best), but in many others the association is poor. In such cases there is generally no shortage of explanations for the apparently unsuitable behavior, but these explanations are, as yet, often just untested hypotheses.

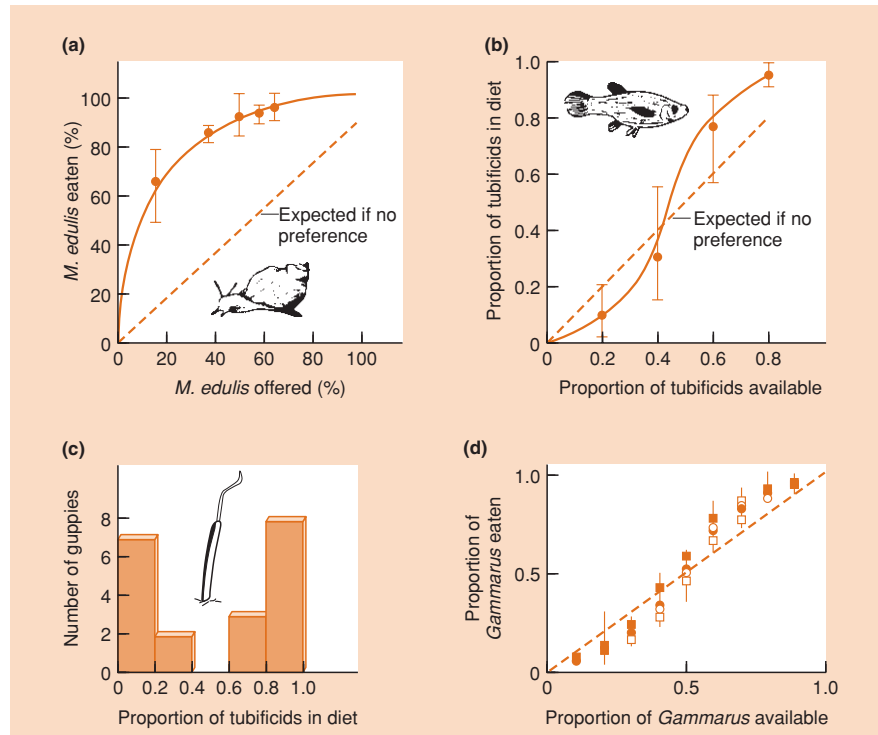


Figure 9.15 Switching. (a) A lack of switching: snails exhibit a consistent preference amongst the mussels *Mytilus edulis* and *M. californianus*, irrespective of their relative abundance (means plus standard errors). (After Murdoch & Stewart-Oaten, 1975.) (b) Switching by guppies fed on tubificids and fruit-flies: they take a disproportionate amount of whichever prey type is the more available (means and total ranges). (After Murdoch *et al.*, 1975.) (c) Preferences shown by the individual guppies in (b) when offered equal amounts of the two prey types: individuals were mostly specialists on one or other type. (d) Switching by sticklebacks fed mixtures of *Gammarus* and *Artemia*: overall they take a disproportionate amount of whichever is more available. However, in the first series of trials, with *Gammarus* availability decreasing (closed symbols), first-day trialists (■) tended to take more *Gammarus* than third-day trialists (●), whereas with *Gammarus* availability increasing, firsts (□) tended to take less *Gammarus* than thirds (○). The effects of learning are apparent. (After Hughes & Croy, 1993.)

9.5.2 Switching

switching involves a preference for food types that are common

The preferences of many consumers are fixed; in other words, they are maintained irrespective of the relative availabilities of alternative food types. But many others switch their preference,

such that food items are eaten disproportionately often when they are common and are disproportionately ignored when they are rare. The two types of preference are contrasted in Figure 9.15. Figure 9.15a shows the fixed preference exhibited by predatory shore snails when they were presented with two species of mussel prey at a range of proportions. The line in Figure 9.15a has been drawn on the assumption that they exhibited the same preference at all proportions. This assumption is clearly justified: irrespective of availability, the predatory snails showed the same marked preference for the thin-shelled, less protected *Mytilus edulis*, which they could exploit more effectively. By contrast,

Figure 9.15b shows what happened when guppies (fish) were offered a choice between fruit-flies and tubificid worms as prey. The guppies clearly switched their preference, and consumed a disproportionate number of the more abundant prey type.

There are a number of situations in which switching can arise. Probably the most common is where different types of prey are found in different microhabitats, and the consumers concentrate on the most profitable microhabitat. This was the case for the guppies in Figure 9.15b: the fruit-flies floated at the water surface whilst the tubificids were found at the bottom. Switching can also occur (Bergelson, 1985) in the following situations:

when might switching arise?

- 1 When there is an increased probability of orientating toward a common prey type, i.e. consumers develop a 'search image' for abundant food (Tinbergen, 1960) and concentrate on their 'image' prey to the relative exclusion of nonimage prey.

- 2 When there is an increased probability of pursuing a common prey type.
- 3 When there is an increased probability of capturing a common prey type.
- 4 When there is an increased efficiency in handling a common prey type.

In each case, increasingly common prey generate increased interest and/or success on the part of the predator, and hence an increased rate of consumption. For instance, switching occurred in the 15-spined stickleback, *Spinachia spinachia*, feeding on the crustaceans *Gammarus* and *Artemia* as alternative prey (Figure 9.15d) as a result of learned improvements in capturing and handling efficiencies, especially of *Gammarus*. Fish were fed *Gammarus* for 7 days, which was then replaced in the diet, in 10% steps, with *Artemia* until the diet was 100% *Artemia*. This diet was then maintained for a further 7 days, when the process was reversed back down to 100% *Gammarus*. Each 'step' itself lasted 3 days, on each of which the fish were tested. The learning process is apparent in Figure 9.15d in the tendency for first-day trialists to be more influenced than third-day trialists by the previous dietary mix.

Interestingly, switching in a population often seems to be a consequence not of individual consumers gradually changing their preference, but of the proportion of specialists changing. Figure 9.15c shows this for the guppies. When the prey types were equally abundant, individual guppies were not generalists – rather, there were approximately equal numbers of fruit-fly and tubificid specialists.

a plant that 'switches'

It may come as a surprise that a plant may show behavior akin to switching. The northern pitcher plant *Sarracenia purpurea* lives in nutrient-poor bogs and fens, circumstances that are thought to favor carnivory in plants. Carnivorous plants such as pitcher plants invest an excess of carbon (captured in photosynthesis) in specialist organs for capturing invertebrate prey (effectively nitrogen-capturing structures). Figure 9.16 shows how relative size of the pitcher keel responded to nitrogen addition to plots in Molly Bog in Vermont, USA. The more nitrogen that was applied, the larger the relative keel size – this corresponds to an increase in size of the noncarnivorous keel of the pitcher and a decrease in size of the prey-catching tube. Thus, with increasing nitrogen levels, the capacity for carnivory decreased while maximum photosynthesis rates increased. In effect, the plants switched effort from nitrogen to carbon capture when more nitrogen was available in their environment.

9.5.3 The optimal foraging approach to diet width

diet width and evolution

Predators and prey have undoubtedly influenced one another's evolution. This can be seen in the distasteful or

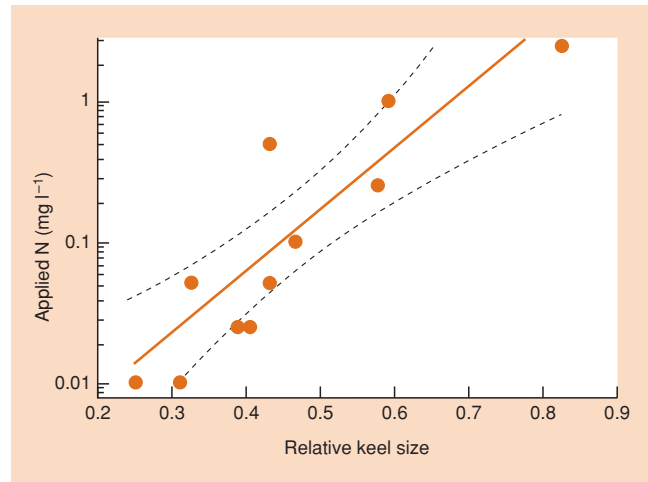


Figure 9.16 The relationship between relative keel size of pitchers of *Sarracenia purpurea* and nitrogen added as aerial spray in plots at Molly Bog, Vermont. Dotted lines indicate 95% confidence intervals. A larger relative keel size corresponds to a reduced investment in organs of prey capture. (After Ellison & Gotelli, 2002.)

poisonous leaves of many plants, in the spines of hedgehogs and in the camouflage coloration of many insect prey; and it can be seen in the stout ovipositors of wood wasps, the multichambered stomachs of cattle and the silent approach and sensory excellence of owls. Such specialization makes it clear, though, that no predator can possibly be capable of consuming all types of prey. Simple design constraints prevent shrews from eating owls (even though shrews are carnivores) and prevent humming-birds from eating seeds.

Even within their constraints, however, most animals consume a narrower range of food types than they are morphologically capable of consuming. In trying to understand what determines a consumer's actual diet within its wide potential range, ecologists have increasingly turned to *optimal foraging theory*. The aim of optimal foraging theory is to predict the foraging strategy to be expected under specified conditions. It generally makes such predictions on the basis of a number of assumptions:

- 1 The foraging behavior that is exhibited by present-day animals is the one that has been favored by natural selection in the past but also most enhances an animal's fitness at present.
- 2 High fitness is achieved by a high net rate of energy intake (i.e. gross energy intake minus the energetic costs of obtaining that energy).
- 3 Experimental animals are observed in an environment to which their foraging behavior is suited, i.e. it is a natural environment very similar to that in which they evolved, or an experimental arena similar in essential respects to the natural environment.

assumptions inherent in optimal foraging theory

These assumptions will not always be justified. First, other aspects of an organism's behavior may influence fitness more than optimal foraging does. For example, there may be such a premium on the avoidance of predators that animals forage at a place and time where the risk from predators is lower, and in consequence gather their food less efficiently than is theoretically possible (see Section 9.5.4). Second, and just as important, for many consumers (particularly herbivores and omnivores) the efficient gathering of energy may be less critical than of some other dietary constituent (e.g. nitrogen), or it may be of prime importance for the forager to consume a mixed and balanced diet. In such cases, the value of existing optimal foraging theory is limited. However, in circumstances where the energy maximization premise can be expected to apply, optimal foraging theory offers a powerful insight into the significance of the foraging 'decisions' that predators make (for reviews see Stephens & Krebs, 1986; Krebs & Kacelnik, 1991; Sih & Christensen, 2001).

theoreticians are
omniscient
mathematicians – the
foragers need not be

Typically, optimal foraging theory makes predictions about foraging behavior based on mathematical models constructed by ecological theoreticians who are omniscient ('all knowing') as far as their model systems are concerned.

The question therefore arises: is it necessary for a real forager to be equally omniscient and mathematical, if it is to adopt the appropriate, optimal strategy? The answer is 'no'. The theory simply says that if there is a forager that in some way (in any way) manages to do the right thing in the right circumstances, then this forager will be favored by natural selection; and if its abilities are inherited, these should spread, in evolutionary time, throughout the population.

Optimal foraging theory does not specify precisely how the forager should make the right decisions, and it does not require the forager to carry out the same calculations as the modeler. Later we consider another group of 'mechanistic' models (see Section 9.6.2) that attempt to show how a forager, given that it is not omniscient, might nevertheless manage to respond by 'rules of thumb' to limited environmental information and thereby exhibit a strategy that is favored by natural selection. But it is optimal foraging theory that predicts the nature of the strategy that should be so favored.

The first paper on optimal foraging theory (MacArthur & Pianka, 1966) sought to understand the determination of diet 'width' (the range of food types eaten by an animal) within a habitat. Subsequently, the model was developed into a more rigorous algebraic form, notably by Charnov (1976a). MacArthur and Pianka argued that to obtain food, any predator must expend time and energy, first in searching for its prey and then in handling it (i.e. pursuing, subduing and consuming it). Whilst searching, a predator is likely to encounter a wide variety of food items. MacArthur and Pianka therefore saw diet width as depending on the responses of predators once they had encountered prey.

Generalists pursue (and may then subdue and consume) a large proportion of the prey types they encounter; specialists continue searching except when they encounter prey of their specifically preferred type.

The 'problem' for any forager is this: if it is a specialist, then it will only pursue profitable prey items, but it may expend a great deal of time and energy searching for them. Whereas if it is a generalist, it will spend relatively little time searching, but it will pursue both more and less profitable types of prey. An optimal forager should balance the pros and cons so as to maximize its overall rate of energy intake. MacArthur and Pianka expressed the problem as follows: given that a predator already includes a certain number of profitable items in its diet, should it expand its diet (and thereby decrease its search time) by including the next most profitable item as well?

to pursue or not
pursue?

We can refer to this 'next most profitable' item as the i th item. E_i/h_i is then the profitability of the item, where E_i is its energy content, and h_i its handling time. In addition, \bar{E}/\bar{h} is the average profitability of the 'present' diet (i.e. one that includes all prey types that are more profitable than i , but does not include prey type i itself), and \bar{s} is the average search time for the present diet. If a predator does pursue a prey item of type i , then its expected rate of energy intake is E_i/h_i . But if it ignores this prey item, whilst pursuing all those that are more profitable, then it can expect to search for a further \bar{s} , following which its expected rate of energy intake is \bar{E}/\bar{h} . The total time spent in this latter case is $\bar{s} + \bar{h}$, and so the overall expected rate of energy intake is $\bar{E}/(\bar{s} + \bar{h})$. The most profitable, optimal strategy for a predator will be to pursue the i th item if, and only if:

$$E_i/h_i \geq \bar{E}/(\bar{s} + \bar{h}). \quad (9.1)$$

In other words, a predator should continue to add increasingly less profitable items to its diet as long as Equation 9.1 is satisfied (i.e. as long as this increases its overall rate of energy intake). This will serve to maximize its overall rate of energy intake, $\bar{E}/(\bar{s} + \bar{h})$.

This optimal diet model leads to a number of predictions.

- 1 Predators with handling times that are typically short compared to their search times should be generalists, because in the short time it takes them to handle a prey item that has already been found, they can barely begin to search for another prey item. (In terms of Equation 9.1: E_i/h_i is large (h_i is small) for a wide range of prey types, whereas $\bar{E}/(\bar{s} + \bar{h})$ is small (\bar{s} is large) even for broad diets.) This prediction seems to be supported by the broad diets of many insectivorous birds feeding in trees and shrubs. Searching is always moderately time consuming, but handling the minute insects takes negligible time and is almost always successful. A bird,

searchers should be
generalists

therefore, has something to gain and virtually nothing to lose by consuming an item once found, and overall profitability is maximized by a broad diet.

handlers should be specialists

2 By contrast, predators with handling times that are long relative to their search times should be specialists. That is, if \bar{s} is always small,

then $\bar{E}/(\bar{s} + \bar{h})$ is similar to \bar{E}/\bar{h} . Thus, maximizing $\bar{E}/(\bar{s} + \bar{h})$ is much the same as maximizing \bar{E}/\bar{h} , which is achieved, clearly, by including only the most profitable items in the diet. For instance, lions live more or less constantly in sight of their prey so that search time is negligible; handling time, on the other hand, and particularly pursuit time, can be long (and very energy consuming). Lions consequently specialize on prey that can be pursued most profitably: the immature, the lame and the old.

specialization should be greater in productive environments

3 Other things being equal, a predator should have a broader diet in an unproductive environment (where prey items are relatively rare and \bar{s} is relatively large) than in a productive environment (where \bar{s} is

smaller). This prediction is broadly supported by the two examples shown in Figure 9.17: in experimental arenas, both bluegill sunfish (*Lepomis macrochirus*) and great tits (*Parus major*) had more specialized diets when prey density was higher. A related result has been reported from predators in their natural setting – brown and black bears (*Ursos arctos* and

U. americanus) feeding on salmon in Bristol Bay in Alaska. When salmon availability was high, bears consumed less biomass per captured fish, targeting energy-rich fish (those that had not spawned) or energy-rich body parts (eggs in females, brain in males). In essence their diet became more specialized when prey were abundant (Gende *et al.*, 2001).

the abundance of unprofitable prey types is irrelevant

4 Equation 9.1 depends on the profitability of the *i*th item (E_i/h_i), depends on the profitabilities of the items already in the diet (\bar{E}/\bar{h}) and depends on the search times for items already in the diet (\bar{s}) and thus on their abundance. But it does not depend on the search time for the *i*th item, s_i . In other words, predators should ignore insufficiently profitable food types irrespective of their abundance. Re-examining the examples in Figure 9.17, we can see that these both refer to cases in which the optimal diet model does indeed predict that the least profitable items should be ignored completely. The foraging behavior was very similar to this prediction, but in both cases the animals consistently took slightly more than expected of the less profitable food types. In fact, this sort of discrepancy has been uncovered repeatedly, and there are a number of reasons why it may occur, which can be summarized crudely by noting that the animals are not omniscient. The optimal diet model, however, does not predict a perfect correspondence between observation and expectation. It predicts the sort of strategy that will be favored by natural selection, and says that the animals that come closest to this

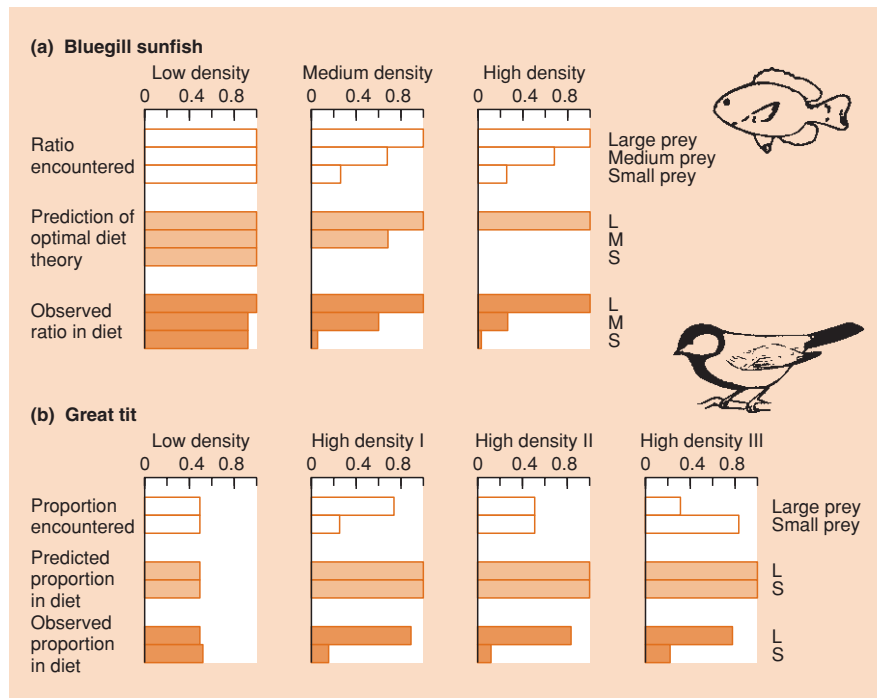


Figure 9.17 Two studies of optimal diet choice that show a clear but limited correspondence with the predictions of Charnov’s (1976a) optimal diet model. Diets are more specialized at high prey densities; but more low profitability items are included than predicted by the theory. (a) Bluegill sunfish preying on different size classes of *Daphnia*: the histograms show ratios of encounter rates with each size class at three different densities, together with the predicted and observed ratios in the diet. (After Werner & Hall, 1974.) (b) Great tits preying on large and small pieces of mealworm. (After Krebs *et al.*, 1977.) The histograms in this case refer to the proportions of the two types of item taken. (After Krebs, 1978.)

strategy will be most favored. From this point of view, the correspondence between data and theory in Figure 9.17 seems much more satisfactory. Sih and Christensen (2001) reviewed 134 studies of optimal diet theory, focusing on the question of what factors might explain the ability of the theory to correctly predict diets. Contrary to their *a priori* prediction, forager groups (invertebrate versus ectothermic vertebrate versus endothermic vertebrate) did not differ in the likelihood of corroborating the theory. Their major conclusion was that while optimal diet theory generally works well for foragers that feed on immobile or relatively immobile prey (leaves, seeds, mealworms, zooplankton relative to fish), it often fails to predict diets of foragers that attack mobile prey (small mammals, fish, zooplankton relative to insect predators). This may be because variations among mobile prey in vulnerability (encounter rate and capture success) are often more important in determining predator diets than are variations in the active choices of predators (Sih & Christensen, 2001).

- 5 Equation 9.1 also provides a context for understanding the narrow specialization of predators that live in intimate association with their prey, especially where an individual predator is linked to an individual prey (e.g. many parasitoids and parasitic herbivores – and many parasites (see Chapter 12)). Since their whole lifestyle and life cycle are finely tuned to those of their prey (or host), handling time (\bar{h}) is low; but this precludes their being finely tuned to other prey species, for which, therefore, handling time is very high. Equation 9.1 will thus only apply within the specialist group, but not to any food item outside it.

On the other hand, polyphagy has definite advantages. Search costs (\bar{s}) are typically low – food is easy to find – and an individual is unlikely to starve because of fluctuations in the abundance of one type of food. In addition, polyphagous consumers can, of course, construct a balanced diet, and maintain this balance by varying preferences to suit altered circumstances, and can avoid consuming large quantities of a toxin produced by one of its food types. These are considerations ignored by Equation 9.1.

coevolution:
predator–prey arms
races?

Overall, then, evolution may broaden or restrict diets. Where prey exert evolutionary pressures demanding specialized morphological or physiological responses from the consumer,

restriction is often taken to extremes. But where consumers feed on items that are individually inaccessible or unpredictable or lacking in certain nutrients, the diet often remains broad. An appealing and much-discussed idea is that particular pairs of predator and prey species have not only evolved but have coevolved. In other words, there has been an evolutionary ‘arms race’, whereby each improvement in predatory ability has been followed by an improvement in the prey’s ability to avoid or resist the predator, which has been followed by a further improvement in

predatory ability, and so on. This may itself be accompanied, on a long-term, evolutionary timescale, by speciation, so that, for example, related species of butterfly are associated with related species of plants – all the species of the Heliconiini feed on members of the Passifloraceae (Ehrlich & Raven, 1964; Futuyma & May, 1992). To the extent that coevolution occurs, it may certainly be an additional force in favor of diet restriction. At present, however, hard evidence for predator–prey or plant–herbivore coevolution is proving difficult to come by (Futuyma & Slatkin, 1983; Futuyma & May, 1992).

There may seem, at first sight, to be a contradiction between the predictions of the optimal diet model and switching. In the latter, a consumer switches from one prey type to another as their relative densities change. But the optimal diet model suggests that the more profitable prey type should always be taken, irrespective of its density or the density of any alternative. Switching is presumed to occur, however, in circumstances to which the optimal diet model does not strictly apply. Specifically, switching often occurs when the different prey types occupy different microhabitats, whereas the optimal diet model predicts behavior within a microhabitat. Moreover, most other cases of switching involve a change in the profitabilities of items of prey as their density changes, whereas in the optimal diet model these are constants. Indeed, in cases of switching, the more abundant prey type is the more profitable, and in such a case the optimal diet model predicts specialization on whichever prey type is more profitable (that is, whichever is more abundant; in other words, switching).

9.5.4 Foraging in a broader context

It is worth stressing that foraging strategies will not always be strategies for simply maximizing feeding efficiency. On the contrary, natural selection will favor foragers that maximize their net benefits, and strategies will therefore often be modified by other, conflicting demands on the individuals concerned. In particular, the need to avoid predators will frequently affect an animal’s foraging behavior.

backswimmers forage suboptimally but avoid being preyed on . . .

This has been shown in work on foraging by nymphs of an aquatic insect predator, the backswimmer *Notonecta hoffmanni* (Sih, 1982). These animals pass through five nymphal instars (with I being the smallest and youngest, and V the oldest), and in the laboratory the first three instars are liable to be preyed upon by adults of the same species, such that the relative risk of predation from adults was:

$$I > II > III > IV = V \cong \text{no risk.}$$

These risks appear to modify the behavior of the nymphs, in that they tend (both in the laboratory and in the field) to avoid the

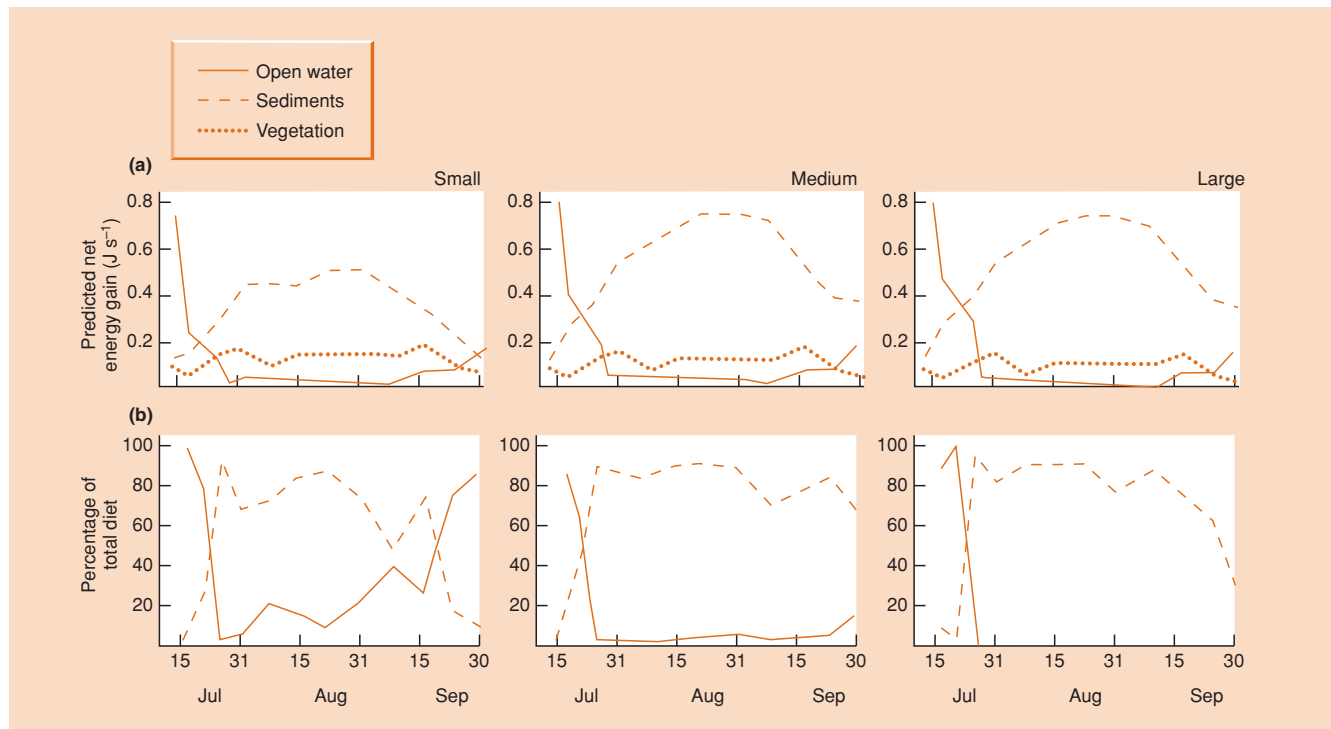


Figure 9.18 Seasonal patterns in (a) the predicted habitat profitabilities (net rate of energy gain) and (b) the actual percentage of the diet originating from each habitat, for three size classes of bluegill sunfish (*Lepomis macrochirus*). Piscivores were absent. (The ‘vegetation’ habitat is omitted from (b) for the sake of clarity – only 8–13% of the diet originated from this habitat for all size classes of fish.) There is good correspondence between the patterns in (a) and (b). (After Werner *et al.*, 1983b.)

central areas of water bodies, where the concentration of adults is greatest. In fact, the relative degree of avoidance was the same as the relative risk of predation from adults:

$$I > II > III > IV = V \cong \text{no avoidance.}$$

Yet these central areas also contain the greatest concentration of prey items for the nymphs, and so, by avoiding predators, nymphs of instars I and II showed a reduction in feeding rate in the presence of adults (although those of instar III did not). The young nymphs displayed a less than maximal feeding rate as a result of their avoidance of predation, but an increased survivorship.

The modifying influence of predators on foraging behavior has also been studied by Werner *et al.* (1983b) working on bluegill sunfish. They estimated the net energy returns from foraging in three contrasting laboratory habitats – in open water, amongst water weeds and on bare sediment – and they examined how prey densities varied in comparable natural habitats in a lake through the seasons. They were then able to predict the time at which the sunfish should switch between different lake habitats so as to maximize their overall net energy returns. In the

... as do certain fish

absence of predators, three sizes of sunfish behaved as predicted (Figure 9.18). But in a further field experiment, this time in the presence of predatory largemouth bass, the small sunfish restricted their foraging to the water weed habitat (Figure 9.19) (Werner *et al.*, 1983a). Here, they were relatively safe from predation, although they could only achieve a markedly submaximal rate of energy intake. By contrast, the larger sunfish are more or less safe from predation by bass, and they continued to forage according to the optimal foraging predictions. In a similar vein, the nymphs of several species of algivorous mayflies largely restrict their feeding to the hours of darkness in streams that contain brown trout, reducing their overall feeding rates but also reducing the risk of predation (Townsend, 2003). In the case of mammals that feed at night, including mice, porcupines and hares, time spent feeding may be reduced in bright moonlight when predation risk is highest (Kie, 1999).

A foraging strategy is an integral part of an animal’s overall pattern of behavior. The strategy is strongly influenced by the selective pressures favoring the maximization of feeding efficiency, but it may also be influenced by other, possibly conflicting demands. It is also worth pointing out one other thing. The places where animals occur, where they are maximally

predation and the realized niche

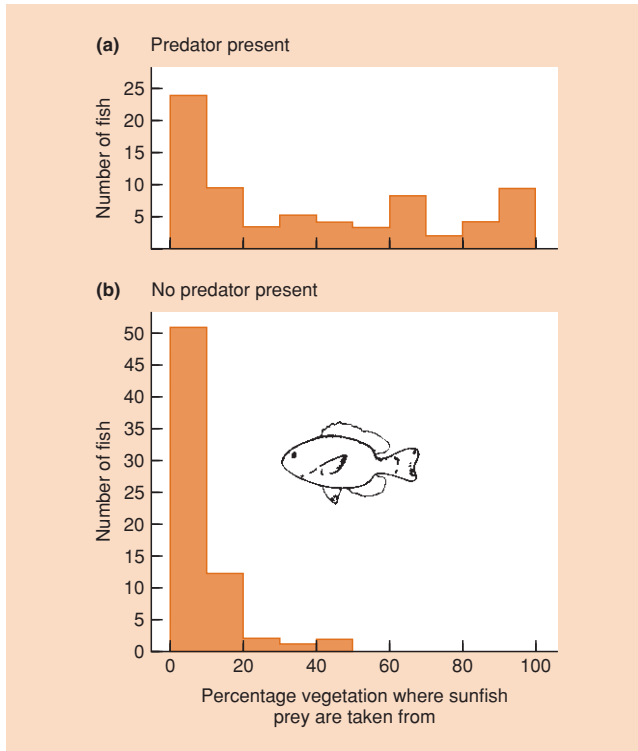


Figure 9.19 (a) In contrast to Figure 9.18 and to (b), when largemouth bass (which prey on small bluegill sunfish) are present many sunfish take prey from areas where the percentage vegetation is high and where they are relatively protected from predation. (After Werner *et al.*, 1983a.)

abundant and where they choose to feed are all key components of their ‘realized niches’. We saw in Chapter 8 that realized niches can be highly constrained by competitors. Here, we see that they can also be highly constrained by predators. This is also seen in the effects of predation by the barn owl (*Tyto alba*) on the foraging behavior of three heteromyid rodents, the Arizona pocket mouse (*Perognathus amplus*), Bailey’s pocket mouse (*P. baileyi*) and Merriam’s kangaroo rat (*Dipodomys merriami*) (Brown *et al.*, 1988). In the presence of owls, all three species moved to microhabitats where they were less at risk from owl predation and where they reduced their foraging activity. However they did so to varying extents, such that the way in which the microhabitat was partitioned between them was quite different in the presence and absence of owls.

9.6 Foraging in a patchy environment

food is patchily distributed

For all consumers, food is distributed patchily. The patches may be natural and discrete physical objects: a bush

laden with berries is a patch for a fruit-eating bird; a leaf covered with aphids is a patch for a predatory ladybird. Alternatively, a ‘patch’ may only exist as an arbitrarily defined area in an apparently uniform environment; for a wading bird feeding on a sandy beach, different 10 m² areas may be thought of as patches that contain different densities of worms. In all cases though, a patch must be defined with a particular consumer in mind. One leaf is an appropriate patch for a ladybird, but for a larger and more active insectivorous bird, 1 m² of canopy or even a whole tree may represent a more appropriate patch.

Ecologists have been particularly interested in patch preferences of consumers where patches vary in the density of food or prey items they contain. There are many examples where predators show an ‘aggregative response’, spending more time in patches containing high densities (because these are the most profitable patches) (Figure 9.20a–d), although such direct density dependence is not always the case (Figure 9.20e). We deal with aggregative responses in more detail in Chapter 10 where their importance in population dynamics will be our focus, and particularly their potential to lend stability to predator–prey dynamics. For now, we concentrate on the behavior that leads to predator aggregation (Section 9.6.1), the optimal foraging approach to patch use (Section 9.6.2) and the distribution patterns that are likely to result when the opposing tendencies of predators to aggregate and to interfere with each other’s foraging are both taken into account (Section 9.6.3).

9.6.1 Behavior that leads to aggregated distributions

There are various types of behavior underlying the aggregative responses of consumers, but they fall into two broad categories: those involved with the location of profitable patches, and the responses of consumers once within a patch. The first category includes all examples in which consumers perceive, at a distance, the existence of heterogeneity in the distribution of their prey.

Within the second category – responses of consumers within patches – there are two main aspects of behavior. The first is a change in the consumer’s pattern of searching after encountering items of food. In particular, there is often a slowing down of movement and an increased rate of turning immediately following the intake of food, both of which lead to the consumer remaining in the vicinity of its last food item (‘area-restricted search’). Alternatively, or in addition, consumers may simply abandon unprofitable patches more rapidly than they abandon profitable ones. Both types of behavior were evident when the carnivorous, net-spinning larva of the caddis-fly *Plectrocnemia conspersa* feeds on chironomid (midge) larvae in a laboratory stream. Caddis in their nets were provided with one prey item at the beginning of the experiment and then fed daily rations of

locating a patch

area-restricted search

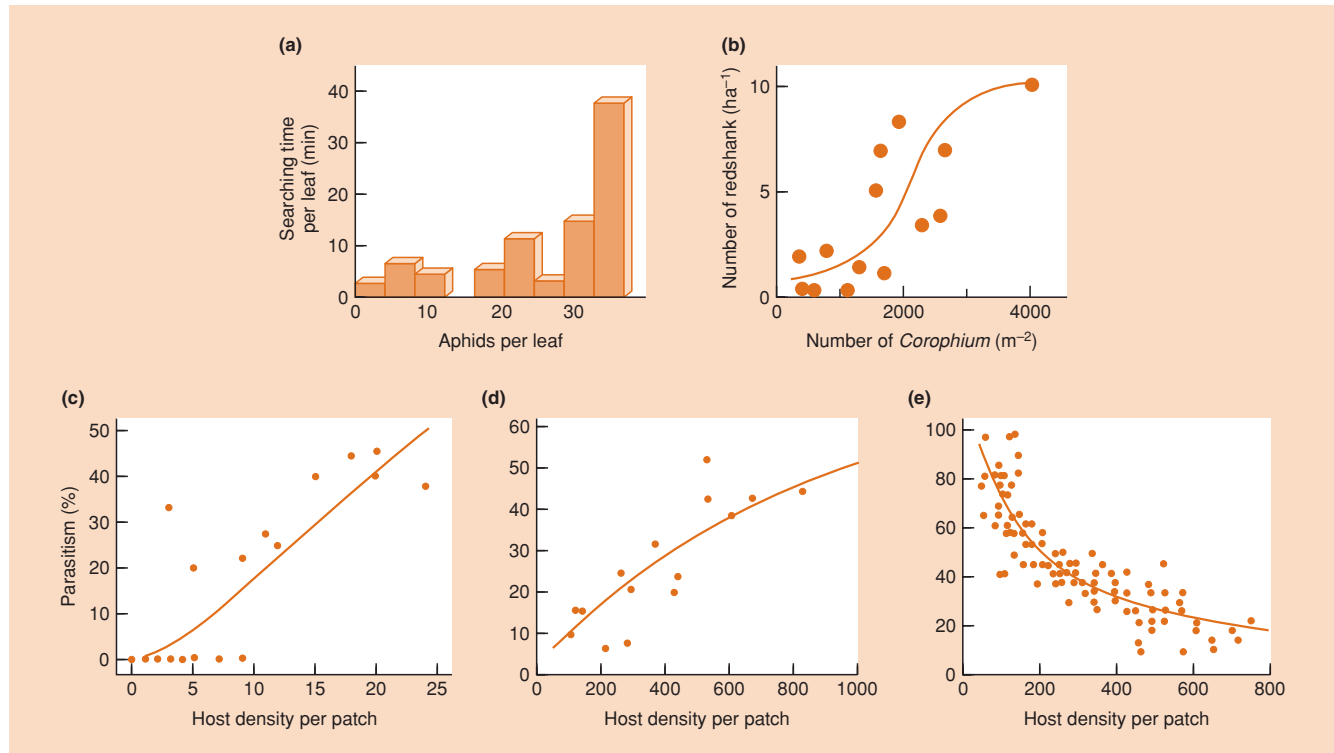


Figure 9.20 Aggregative responses: (a) coccinellid larvae (*Coccinella septempunctata*) spend more time on leaves with high densities of their aphid prey (*Brevicoryne brassicae*) (after Hassell & May, 1974); (b) redshank (*Tringa totanus*) aggregate in patches with higher densities of their amphipod prey (*Corophium volutator*) (after Goss-Custard, 1970); (c) direct density dependence when the parasitoid *Delia radicum* attacks *Trybliographa rapae*; and (d) direct density dependence when the parasitoid *Aspidiotiphagus citrinus* attacks *Fiorinia externa*. (e) But direct density dependence is not always the case: inverse density dependence when the parasitoid *Ooencyrtus kuwanai* attacks *Lymantria dispar*. ((c–e) after Pacala & Hassall, 1991.)

zero, one or three prey. The tendency to abandon the net was lowest at the higher feeding rates (Townsend & Hildrew, 1980). *Plectrocnemia*'s behavior in relation to prey patches also has an element of area-restricted search: the likelihood that it will spin a net in the first place depends on whether it happens to encounter a food item (which it can consume even without a net) (Figure 9.21a). Overall, therefore, a net is more likely to be constructed, and less likely to be abandoned, in a rich patch. These two behaviors account for a directly density-dependent aggregative response in the natural stream environment observed for much of the year (Figure 9.21b).

The difference in the rates of abandonment of patches of high and low profitability can be achieved in a number of ways, but two are especially easy to envisage. A consumer might leave a patch when its feeding rate drops below a threshold level, or a consumer might have a giving-up time – it might abandon a patch whenever a particular time interval passes without the successful capture of food. Whichever mechanism is used, or indeed if the consumer simply uses area-restricted search, the consequences will be the same:

thresholds and giving-up times

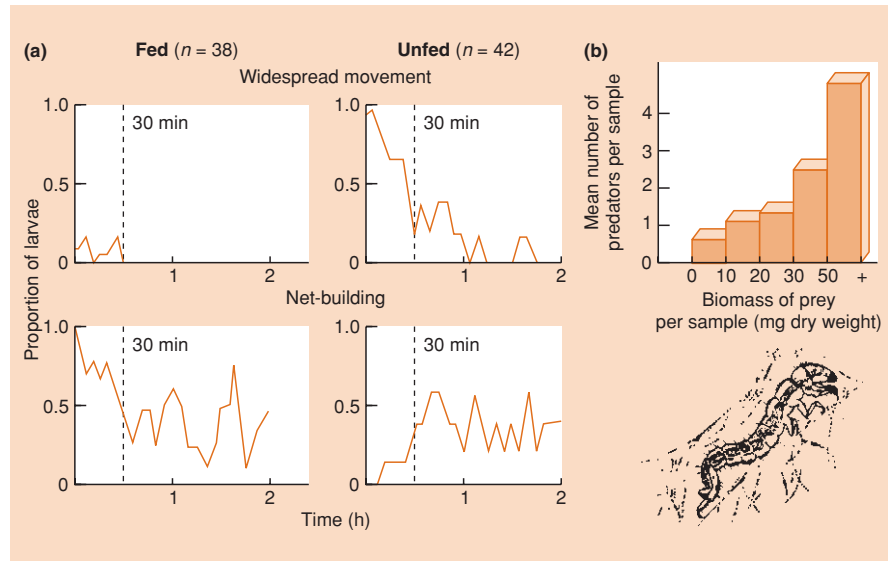
individuals will spend longer in more profitable patches, and these patches will therefore generally contain more consumers.

9.6.2 Optimal foraging approach to patch use

The advantages to a consumer of spending more time in higher profitability patches are easy to see. However, the detailed allocation of time to different patches is a subtle problem, since it depends on the precise differentials in profitability, the average profitability of the environment as a whole, the distance between the patches, and so on. The problem has been a particular focus of attention for optimal foraging theory. In particular, a great deal of interest has been directed at the very common situation in which foragers themselves deplete the resources of a patch, causing its profitability to decline with time. Amongst the many examples of this are insectivorous insects removing prey from a leaf, and bees consuming nectar from a flower.

Charnov (1976b) and Parker and Stuart (1976) produced similar models to predict the behavior of an optimal forager in such situations. They found that the optimal stay-time in a patch

Figure 9.21 (a) On arrival in a patch, fifth-instar *Plectrocnemia conspersa* larvae that encounter and eat a chironomid prey item at the beginning of the experiment ('fed') quickly cease wandering and commence net-building. Predators that fail to encounter a prey item ('unfed') exhibit much more widespread movement during the first 30 min of the experiment, and are significantly more likely to move out of the patch. (b) Directly density-dependent aggregative response of fifth-instar larvae in a natural environment expressed as mean number of predators against combined biomass of chironomid and stonefly prey per 0.0625 m² sample of streambed ($n = 40$). (After Hildrew & Townsend, 1980; Townsend & Hildrew, 1980.)



should be defined in terms of the rate of energy extraction experienced by the forager at the moment it leaves a patch (the 'marginal value' of the patch). Charnov called the results the 'marginal value theorem'. The models were formulated mathematically, but their salient features are shown in graphic form in Figure 9.22.

The primary assumption of the model is that an optimal forager will maximize its overall intake of a resource (usually energy) during a bout of foraging, taken as a whole. Energy will, in fact, be extracted in bursts because the food is distributed patchily; the forager will sometimes move between patches, during which time its intake of energy will be zero. But once in a patch, the forager will extract energy in a manner described by the curves in Figure 9.22a. Its initial rate of extraction will be high, but as time progresses and the resources are depleted, the rate of extraction will steadily decline. Of course, the rate will itself depend on the initial contents of the patch and on the forager's efficiency and motivation (Figure 9.22a).

when should a forager leave a patch that it is depleting?

The problem under consideration is this: at what point should a forager leave a patch? If it left all patches immediately after reaching them, then it would spend most of its time travel-

ing between patches, and its overall rate of intake would be low. If it stayed in all patches for considerable lengths of time, then it would spend little time traveling, but it would spend extended periods in depleted patches, and its overall rate of intake would again be low. Some intermediate stay-time is therefore optimal. In addition, though, the optimal stay-time must clearly be greater for profitable patches than for unprofitable ones, and it must depend on the profitability of the environment as a whole.

Consider, in particular, the forager in Figure 9.22b. It is foraging in an environment where food is distributed patchily and

where some patches are more valuable than others. The average traveling time between patches is t_t . This is therefore the length of time the forager can expect to spend on average after leaving one patch before it finds another. The forager in Figure 9.22b has arrived at an average patch for its particular environment, and it therefore follows an average extraction curve. In order to forage optimally it must maximize its rate of energy intake not merely for its period in the patch, but for the whole period since its departure from the last patch (i.e. for the period $t_t + s$, where s is the stay-time in the patch).

If it leaves the patch rapidly then this period will be short ($t_t + s_{\text{short}}$ in Figure 9.22b). But by the same token, little energy will be extracted (E_{short}). The rate of extraction (for the whole period $t_t + s$) will be given by the slope of the line OS (i.e. $E_{\text{short}}/(t_t + s_{\text{short}})$). On the other hand, if the forager remains for a long period (s_{long}) then far more energy will be extracted (E_{long}); but, the overall rate of extraction (the slope of OL) will be little changed. To maximize the rate of extraction over the period $t_t + s$, it is necessary to maximize the slope of the line from O to the extraction curve. This is achieved simply by making the line a tangent to the curve (OP in Figure 9.22b). No line from O to the curve can be steeper, and the stay-time associated with it is therefore optimal (s_{opt}).

The optimal solution for the forager in Figure 9.22b, therefore, is to leave that patch when its extraction rate is equal to (tangential to) the slope of OP, i.e. it should leave at point P. In fact, Charnov, and Parker and Stuart, found that the optimal solution for the forager is to leave all patches, irrespective of their profitability, at the same extraction rate (i.e. the same 'marginal value'). This extraction rate is given by the slope of the tangent to the average extraction curve (e.g. in Figure 9.22b), and it is therefore the maximum average overall rate for that environment as a whole.

how to maximize overall energy intake

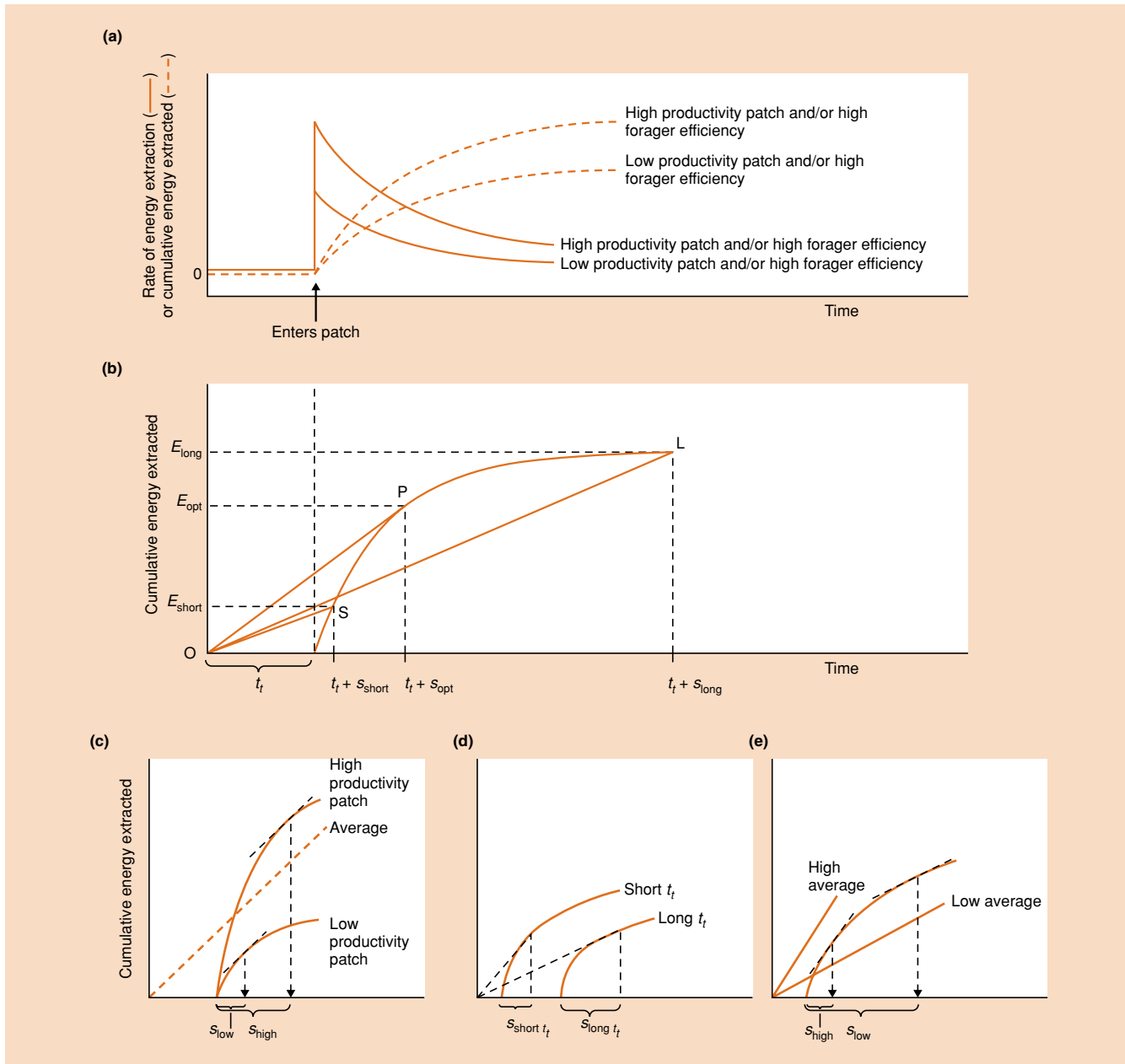


Figure 9.22 The marginal value theorem. (a) When a forager enters a patch, its rate of energy extraction is initially high (especially in a highly productive patch or where the forager has a high foraging efficiency), but this rate declines with time as the patch becomes depleted. The cumulative energy intake approaches an asymptote. (b) The options for a forager. The solid colored curve is cumulative energy extracted from an average patch, and t_i is the average traveling time between patches. The rate of energy extraction (which should be maximized) is energy extracted divided by total time, i.e. the slope of a straight line from the origin to the curve. Short stays in the patch (slope = $E_{\text{short}}/(t_i + s_{\text{short}})$) and long stays (slope = $E_{\text{long}}/(t_i + s_{\text{long}})$) both have lower rates of energy extraction (shallower slopes) than a stay (s_{opt}) which leads to a line just tangential to the curve. s_{opt} is therefore the optimum stay-time, giving the maximum overall rate of energy extraction. All patches should be abandoned at the *same* rate of energy extraction (the slope of the line OP). (c) Low productivity patches should be abandoned after shorter stays than high productivity patches. (d) Patches should be abandoned more quickly when traveling time is short than when it is long. (e) Patches should be abandoned more quickly when the average overall productivity is high than when it is low.

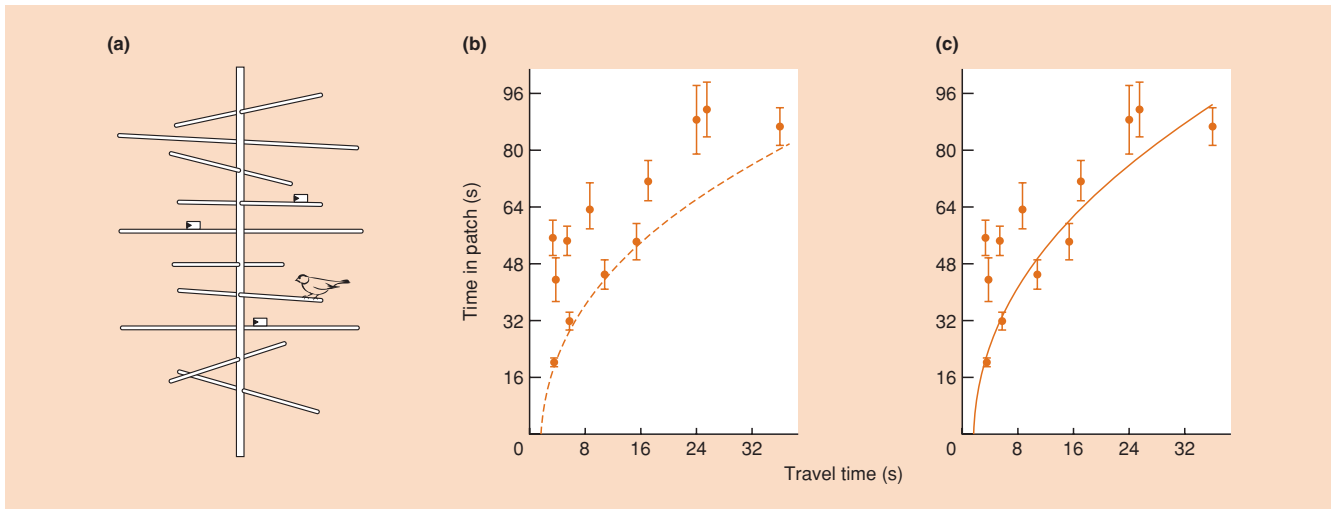


Figure 9.23 (a) An experimental ‘tree’ for great tits, with three patches. (b) Predicted optimal time in a patch plotted against traveling time (-----), together with the observed mean points (\pm SE) for six birds, each in two environments. (c) The same data points, and the predicted time taking into account the energetic costs of traveling between patches. (After Cowie, 1977; from Krebs, 1978.)

predictions of
the marginal
value theorem . . .

The model therefore confirms that the optimal stay-time should be greater in more productive patches than in less productive patches (Figure 9.22c). Moreover, for the least productive patches (where the extraction rate is never as high as OP) the stay-time should be zero. The model also predicts that all patches should be depleted such that the final extraction rate from each is the same (i.e. the ‘marginal value’ of each is the same); and it predicts that stay-times should be longer in environments where the traveling time between patches is longer (Figure 9.22d) and that stay-times should be longer where the environment as a whole is less profitable (Figure 9.22e).

. . . supported by
some experiments

Encouragingly, there is evidence from a number of cases that lends support to the marginal value theorem. In one of the first tests of the theory, Cowie (1977) considered the prediction set out in Figure 9.22d: that a forager should spend longer in each patch when the traveling time is longer. He used captive great tits in a large indoor aviary, and got the birds to forage for small pieces of mealworm hidden in sawdust-filled plastic cups – the cups were ‘patches’. All patches on all occasions contained the same number of prey, but traveling time was manipulated by covering the cups with cardboard lids that varied in their tightness and therefore varied in the time needed to prize them off. Birds foraged alone, and Cowie used six in all, subjecting each to two habitats. One of these habitats always had longer traveling times (tighter lids) than the other. For each bird in each habitat Cowie measured the average traveling time and the curve of cumulative food intake within a patch. He then used the marginal value theorem to predict the optimal stay-time in habitats with different traveling

times, and compared these predictions with the stay-times he actually observed. As Figure 9.23 shows, the correspondence was quite close. It was closer still when he took account of the fact that there was a net loss of energy when the birds were traveling between patches.

Predictions of the marginal value theorem have also been examined through the behavior of the egg parasitoid, *Anaphes victus*, attacking the beetle *Listronotus oregonensis* in a laboratory setting (Boivin *et al.*, 2004). Patches differed in quality by virtue of the varying proportions of hosts already parasitized at the start of the experiment, and in line with the theorem’s predictions, parasitoids stayed longer in the more profitable patches (Figure 9.24a). However, contrary to a further prediction, the marginal rate of fitness gain (the rate of progeny production in the final 10 min before leaving a patch) was greatest in the initially most profitable patches (Figure 9.24b).

As was the case with optimal diet theory, the risk of being preyed upon can be expected to modify the predicted outcomes of optimal patch use. With this in mind, Morris and Davidson (2000) compared the giving-up food extraction rates of white-footed mice (*Peromyscus leucopus*) in a forest habitat (where predation risk is low) and a forest-edge habitat (where predation risk is high). They provided ‘patches’ (containers) with 4 g of millet grain in 11 foraging sites in the two habitat types, and in both habitat types some sites were in relatively open situations and others were beneath shrubs. They then monitored the grain remaining at the time that the patches were abandoned on two separate days. Their results (Figure 9.25) supported the predictions that mice should abandon patches at

optimal patch
use predictions are
modified when there
is a risk of being
preyed upon

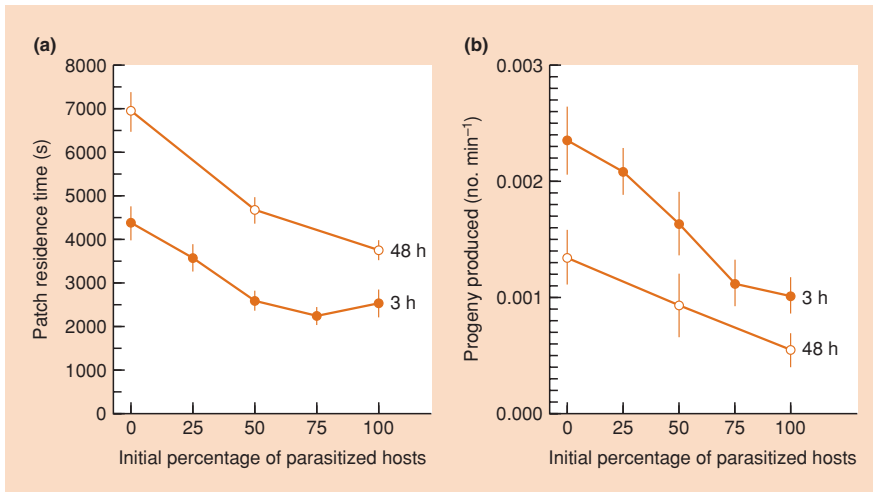


Figure 9.24 (a) When the parasitoid *Anaphes victus* attacked the beetle *Listronotus oregonensis* in patches of 16 hosts, a varying percentage of which had already been parasitized, parasitoids remained longer in the more profitable patches: those with the smaller percentage of parasitized hosts. (b) However, the marginal gain rate in fitness – the number of progeny produced per minute in the final 10 min before leaving a patch – was greatest in the initially most profitable patches. Bars represent standard errors. (After Boivin *et al.*, 2004.)

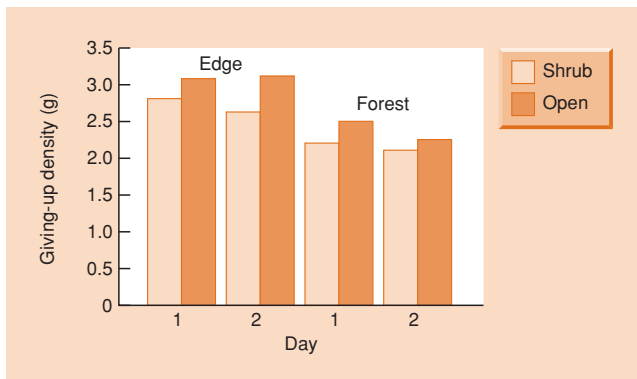


Figure 9.25 The mass of millet grain remaining (giving-up density, g) was higher in patches in the open (riskier) than in paired patches located under shrubs (safer), and was higher in forest-edge habitat (higher predation) than in forest (lower predation). (After Morris & Davidson, 2000.)

higher harvest rates in vulnerable edge habitats than in safe forest habitats, particularly in open situations (where predation risk is highest in each habitat).

predicted and observed behaviors do not correspond perfectly

A much fuller review of tests of the marginal value theorem is provided, for example, by Krebs and Kacelnik (1991). The picture this conveys is one of encouraging but not perfect correspondence – much like the balance

of the results presented here. The main reason for the imperfection is that the animals, unlike the modelers, are not omniscient. As was clear in the case of the white-footed mice, they may need to spend time doing things other than foraging (e.g. avoiding predators). Foragers may also need to spend time learning about and sampling their environment, and are none the less likely to proceed in their foraging with imperfect information about the

distribution of their hosts. For the parasitoids in Figure 9.24, for example, Boivin *et al.* (2004) suggest that they seem to base their assessment of overall habitat quality on the quality of the first patch they encounter; that is, they ‘learn’ but their learned assessment may still be wrong. Such a strategy would be adaptive, though, if there was considerable variation in quality between generations (so that each generation had to learn anew), but little variation in quality between patches within a generation (so that the first patch encountered was a fair indication of quality overall).

Nevertheless, in spite of their limited information, animals seem often to come remarkably close to the predicted strategy. Ollason (1980) developed a mechanistic model to account for this in the great tits studied by Cowie. Ollason’s is a memory model. It assumes that an animal has a ‘remembrance of past food’, which Ollason likens to a bath of water without a plug. Fresh remembrance flows in every time the animal feeds. But remembrance is also draining away continuously. The rate of input depends on the animal’s feeding efficiency and the productivity of the current feeding area. The rate of outflow depends on the animal’s ability to memorize and the amount of remembrance. Remembrance drains away quickly, for example, when the amount is large (high water level) or the memorizing ability is poor (tall, narrow bath). Ollason’s model simply proposes that an animal should stay in a patch until remembrance ceases to rise; an animal should leave a patch when its rate of input from feeding is slower than its rate of declining remembrance.

mechanistic models of optimal foraging

An animal foraging consistently with Ollason’s model behaves in a way very similar to that predicted by the marginal value theorem. This is shown for the case of Cowie’s great tits in Figure 9.26. As Ollason himself remarks, this shows that to forage in a patchy environment in a way that approximates closely to optimality, an animal need not be omniscient, it does not need to sample and it does not need to perform numerical analyses to find the maxima of functions of many variables: all it

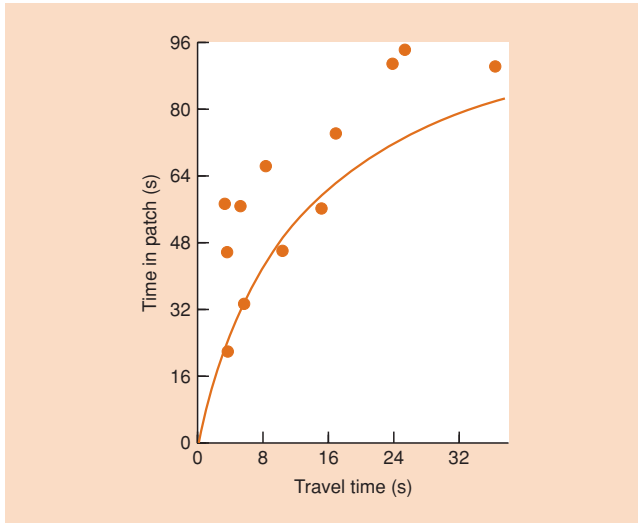


Figure 9.26 Cowie's (1977) great tit data (see Figure 9.23) compared to the predictions of Ollason's (1980) mechanistic memory model.

needs to do is to remember, and to leave each patch if it is not feeding as fast as it remembers doing. As Krebs and Davies (1993) point out, this is no more surprising than the observation that the same birds can fly without any formal qualification in aerodynamics.

Mechanistic models have also been developed and tested for a range of patterns of parasitoid attack (like that in Figure 9.24) (see Vos *et al.*, 1998; Boivin *et al.*, 2004). These highlight the important distinction between 'rule of thumb' behavior, where animals follow innate and unvarying rules, and learned behavior, where rules are subject to modification in the light of the forager's immediate experience. The weight of evidence suggests that learning plays at least some role in most foragers' decisions. There is an important distinction, too, between 'incremental' and 'decremental' behavior. With incremental behavior each successful attack in a patch increases the forager's chance of staying there. This is likely to be adaptive when there is considerable variation in quality between patches, because it encourages longer stay-times in better quality patches. With 'decremental' behavior each successful attack in a patch *decreases* the forager's chance of staying there. This is likely to be adaptive when all patches are of approximately the same quality, because it encourages foragers to leave depleted patches.

Thus, Ollason's model for great tits incorporated rule of thumb, incremental behavior. Boivin *et al.*, on the other hand, found their parasitoids to be exhibiting learned, decremental behavior: a parasitoid attacking a healthy host, for example, was 1.43 times more likely to leave a patch thereafter, and one rejecting a host that had already been attacked was 1.11 times more likely to leave. Vos *et al.* (1998), by contrast, found incremental behavior when the parasitoid *Cotesia glomerata* attacked its

butterfly larva host, *Pieris brassicae*: each successful encounter increased its tendency to remain in a patch. For both the great tit and parasitoids, therefore, optimal foraging and mechanistic models are seen to be compatible and complementary in explaining how a predator has achieved its observed foraging pattern, and why that pattern has been favored by natural selection.

Finally, the principles of optimal foraging are also being applied to investigations of the foraging strategies of plants for nutrients (reviewed by Hutchings & de Kroon, 1994). When does it pay to produce long stolons moving rapidly from patch to patch? When does it pay to concentrate root growth within a limited volume, foraging from a patch until it is close to depletion? Certainly, it is good to see such intellectual cross-fertilization across the taxonomic divide.

optimal foraging
in plants

9.6.3 Ideal free and related distributions: aggregation and interference

We can see, then, that consumers tend to aggregate in profitable patches where their expected rate of food consumption is highest. Yet we might also expect that consumers will compete and interfere with one another (discussed further in Chapter 10), thereby reducing their per capita consumption rate. It follows from this that patches that are initially most profitable become immediately less profitable because they attract most consumers. We might therefore expect the consumers to redistribute themselves, and it is perhaps not surprising that the observed patterns of predator distributions across prey patches vary substantially from case to case. But can we make some sense of this variation in pattern?

the ideal free
distribution . . .

In an early attempt to do so, it was proposed that if a consumer forages optimally, the process of redistribution will continue until the profitabilities of all patches are equal (Fretwell & Lucas, 1970; Parker, 1970). This will happen because as long as there are dissimilar profitabilities, consumers should leave less profitable patches and be attracted to more profitable ones. Fretwell and Lucas called the consequent distribution the ideal free distribution: the consumers are 'ideal' in their judgement of profitability, and 'free' to move from patch to patch. Consumers were also assumed to be equal. Hence, with an ideal free distribution, because all patches come to have the same profitability, all consumers have the same consumption rate. There are some simple cases where consumers appear to conform to an ideal free distribution insofar as they distribute themselves in proportion to the profitabilities of different patches (e.g. Figure 9.27a), but even in such cases one of the underlying assumptions is likely to have been violated (e.g. Figure 9.27b – all consumers are not equal).

. . . is a balance
between attractive
and repellant forces

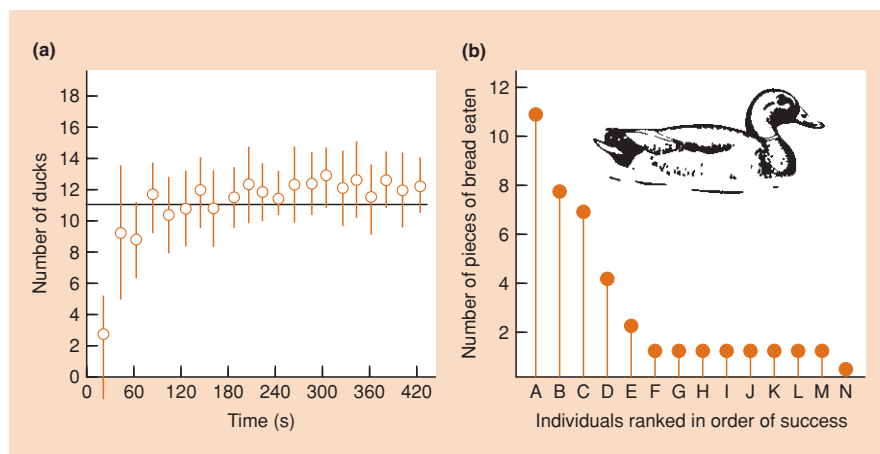


Figure 9.27 (a) When 33 ducks were fed pieces of bread at two stations around a pond (with a profitability ratio of 2 : 1), the number of ducks at the poorer station, shown here, rapidly approached one-third of the total, in apparent conformity with the predictions of ideal free theory. (b) However, contrary to the assumptions and other predictions of simple theory, the ducks were not all equal. (After Harper, 1982; from Milinski & Parker, 1991.)

incorporating a range of interference coefficients

The early ideas have been much modified taking account, for example, of unequal competitors (see Milinski & Parker, 1991; Tregenza, 1995, for reviews). In particular, ideal free theory was put in a more ecological context by Sutherland (1983) when he explicitly incorporated predator handling times and mutual interference amongst the predators. He found that predators should be distributed such that the proportion of predators in site i , p_i , is related to the proportion of prey (or hosts) in site i , h_i , by the equation:

$$p_i = k (h_i^{1/m}) \quad (9.2)$$

where m is the coefficient of interference, and k is a 'normalizing constant' such that the proportions, p_i , add up to 1. It is now possible to see how the patch to patch distribution of predators might be determined jointly by interference and the selection by the predators of intrinsically profitable patches.

If there is no interference amongst the predators, then $m = 0$. All should exploit only the patch with the highest prey density (Figure 9.28), leaving lower density patches devoid of predators.

If there is a small or moderate amount of interference (i.e. $m > 0$, but $m < 1$ – a biologically realistic range), then high-density prey patches should still attract a disproportionate number of predators (Figure 9.28). In other words, there should be an aggregative response by the predators, which is not only directly density dependent, but actually accelerates with increasing prey density in a patch. Hence, the prey's risk of predation might itself be expected to be directly density dependent: the greatest risk of predation in the highest prey density patches (like the examples in Figure 9.20c, d).

With a little more interference ($m \approx 1$) the proportion of the predator population in a patch should still increase with the proportion of prey, but now it should do so more or less linearly rather than accelerating, such that the ratio of predators : prey is roughly the same in all patches (Figure 9.28, and, for example,

Figure 9.20a). Here, therefore, the risk of predation might be expected to be the same in all patches, and hence independent of prey density.

Finally, with a great deal of interference ($m > 1$) the highest density prey patches should have the lowest ratio of predators : prey (Figure 9.28). The risk of predation might therefore be expected to be greatest in the lowest prey density patches, and hence inversely density dependent (like the data in Figure 9.20e).

It is clear, therefore, that the range of patterns amongst the data in Figure 9.20 reflects a shifting balance between the forces of attraction and of repulsion. Predators are attracted to highly profitable patches; but they are repelled by the presence of other predators that have been attracted in the same way.

This description, however, of the relationship between the distribution of predators and the distribution of predation risk has been peppered with 'might be expected to's. The

pseudo-interference

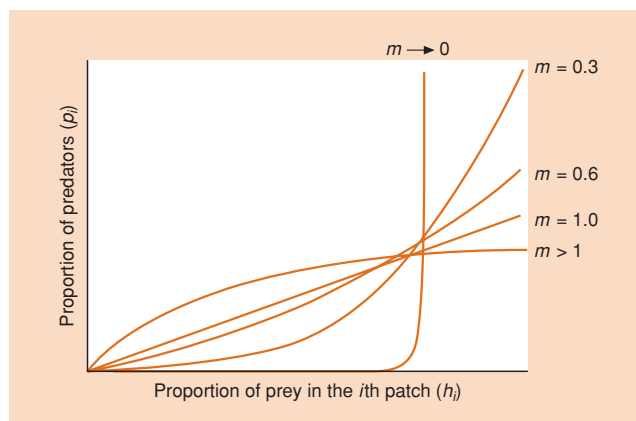


Figure 9.28 The effect of the interference coefficient, m , on the expected distribution of predators amongst patches of prey varying in the proportion of the total prey population they contain (and hence, in their 'intrinsic' profitability). (After Sutherland, 1983.)

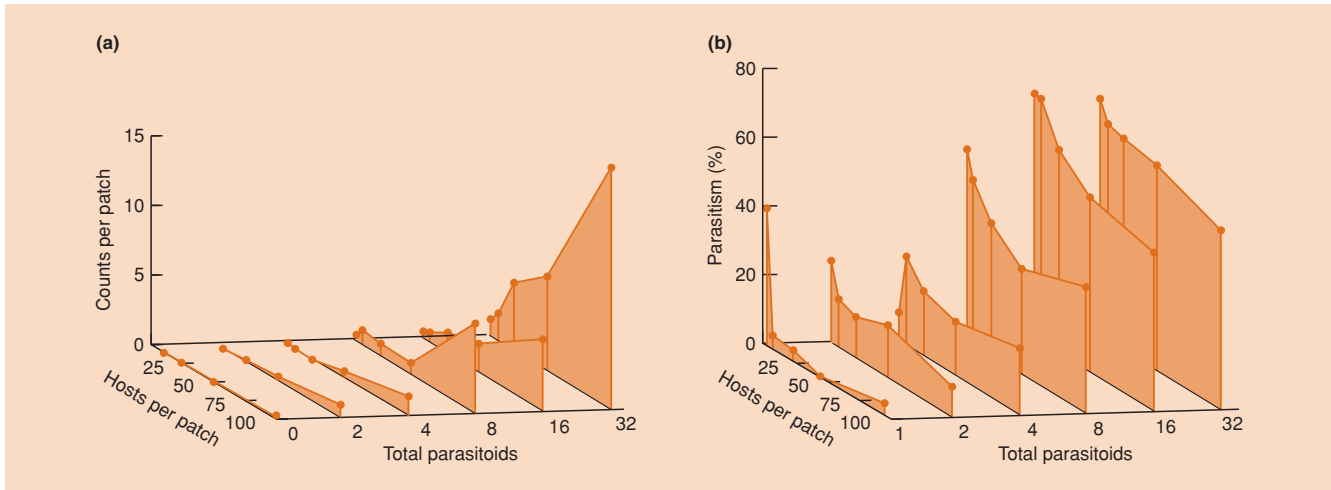


Figure 9.29 (a) The aggregative response of the egg parasitoid, *Trichogramma pretiosum*, which aggregates on patches with high densities of its host *Plodia interpunctella*. (b) The resultant distribution of ill effects: hosts on high-density patches are least likely to be parasitized. (After Hassell, 1982.)

reason is that the relationship also depends on a range of factors not so far considered. For example, Figure 9.29 shows a case where the parasitoid *Trichogramma pretiosum* aggregates in high-density patches of its moth host, but the risk of parasitism to the moth is greatest in low-density host patches. The explanation probably lies in time wasted by parasitoids in high-density host patches, dealing with already parasitized hosts that may still attract parasitoids because they are not physically removed from a patch (unlike preyed-upon prey) (Morrison & Strong, 1981; Hassell, 1982). Thus, earlier parasitoids in a patch may interfere indirectly with later arrivals, in that the previous presence of a parasitoid in a patch may reduce the effective rate at which later arrivals attack unparasitized hosts. This effect has been termed 'pseudo-interference' (Free *et al.*, 1977); its potentially important effects on population dynamics are discussed in Chapter 10.

learning and migration

Expected patterns are modified further still if we incorporate learning by the predators, or the costs of migration between patches (Bernstein *et al.*, 1988, 1991). With a realistic value of m ($= 0.3$), the aggregative response of predators is directly density dependent (as expected) as long as the predators' learning response is strong relative to the rate at which they can deplete patches. But if their learning response is weak, predators may be unable to track the changes in prey density that result from patch depletion. Their distribution will then drift to one that is independent of the density of prey.

Similarly, when the cost of migration is low, the predators' aggregative response remains directly density dependent (with $m = 0.3$) (Figure 9.30a). When the cost of migration is increased, however, it still pays predators in the poorest patches to move, but for others the costs of migration can outweigh the potential gains of moving. For these, the distribution across prey patches

is random. This results in inverse density dependence in mortality rate between intermediate and good patches, and in a 'domed' relationship overall (Figure 9.30b). When the cost of migration is very high, it does not pay predators to move whatever patch they are in – mortality is inversely density dependent across all patches (Figure 9.30c).

Clearly, there is no shortage of potential causes for the wide range of types of distributions of predators, and of mortality rates, across prey patches (see Figures 9.20 and 9.29). Their consequences, in terms of population dynamics, are one of the topics dealt with in the chapter that follows. This highlights the crucial importance of forging links between behavioral and population ecology.

Summary

Predation is the consumption of one organism by another, in which the prey is alive when the predator first attacks it. There are two main ways in which predators can be classified. The first is 'taxonomic' – carnivores consume animals, herbivores consume plants, etc. – and the second is 'functional', in which true predators, grazers, parasitoids and parasites are distinguished.

The effects of herbivory on a plant depend on which herbivores are involved, which plant parts are affected, and the timing of attack relative to the plant's development. Leaf-biting, sap-sucking, mining, flower and fruit damage and root pruning can be expected to differ in the effect they have on the plant. Because the plant usually remains alive in the short term, the effects of herbivory are also crucially dependent on the response of the plant. The evolutionary selection pressure exerted by herbivores has led to a variety of plant physical and chemical defenses that

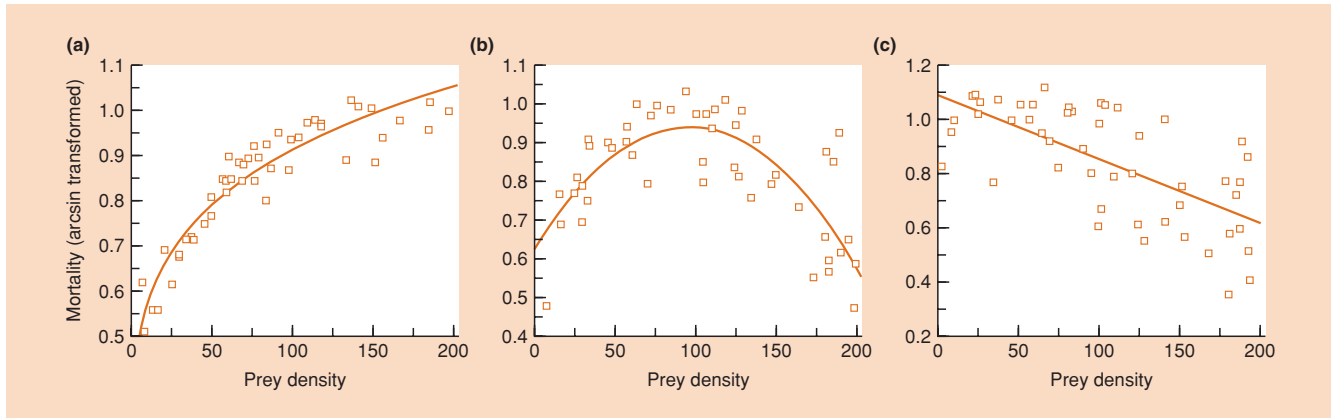


Figure 9.30 The effect of a cost to migration in predators distributing themselves across prey patches in a simulation model. The interference coefficient, m , is 0.3 and would lead to direct density dependence in the absence of a migration cost. (a) Low migration cost: direct density dependence is maintained. (b) Intermediate cost: a ‘domed’ relationship. (c) High cost: inverse density dependence. (After Bernstein *et al.*, 1991.)

resist attack. These may be present and effective continuously (constitutive defense) or increased production may be induced by attack (inducible defense). It is not straightforward to determine whether the supposed ‘defenses’ actually have measurable, negative effects on the herbivore and positive consequences for the plant, especially after the costs of mounting the response have been taken into account. We discuss the difficulties of revealing such effects and review the relationships between herbivory and plant survival and fecundity.

More generally, the immediate effect of predation on a population of prey is not always predictably harmful, first because the individuals that are killed are not always a random sample (and may be those with the lowest potential to contribute to the population’s future) and second because of compensatory changes in the growth, survival or reproduction of the surviving prey (especially via reduced competition for a limiting resource). From the predator’s point of view, an increase in the amount of food consumed can be expected to lead to increased rates of growth, development and birth, and decreased rates of mortality. However, there are a number of factors that complicate this simple relationship between consumption rate and consumer benefit.

Consumers can be classified on a continuum from monophagy (feeding on a single prey type) to polyphagy (many prey types). The preferences of many consumers are fixed – they are main-

tained irrespective of the relative availabilities of alternative food types. But many others switch their preference, such that food items are eaten disproportionately often when they are common. A mixed diet may be favored first because each food type contains a different undesirable toxic chemical. More generally, a generalist strategy would be favored if a consumer has more to gain than lose in accepting low-quality items, once encountered, rather than ignoring them and continuing to search. We discuss this in the context of optimal diet theory, the aim of which is to predict the foraging strategy to be expected under specified conditions.

Food is generally distributed patchily and ecologists have been particularly interested in patch preferences of consumers where patches vary in the density of food or prey items they contain. We describe the behaviors that lead to aggregated distributions and the nature of the distribution patterns that result. The advantages to a consumer of spending more time in higher profitability patches are easy to see. However, the detailed allocation of time to different patches is a subtle problem, depending on the precise differentials in profitability, the average profitability of the environment as a whole, the distance between patches, and so on. This is the domain of the theory of optimal patch use. The predictions of both optimal foraging and optimal patch use theory have to be modified when there is a simultaneous risk of a consumer being preyed upon.

Chapter 10

The Population Dynamics of Predation



10.1 Introduction: patterns of abundance and the need for their explanation

We turn now to the effects of predation on the population dynamics of the predator and its prey, where even a limited survey of the data reveals a varied array of patterns. There are certainly cases where predation has a profoundly detrimental effect on a prey population. For example, the ‘vedalia’ ladybird beetle (*Rodolia cardinalis*) is famous for having virtually eradicated the cottony cushion-scale insect (*Icerya purchasi*), a pest that threatened the California citrus industry in the late 1880s (see Section 15.2.5). On the other hand, there are many cases where predators and herbivores have no apparent effect on their prey’s dynamics or abundance. For example, the weevil *Apion ulicis* has been introduced into many parts of the world in an attempt to control the abundance of gorse bushes (*Ulex europaeus*), and it has often become well established. The situation in Chile, however, is fairly typical, where, despite eating on average around half, and sometimes up to 94%, of the seeds produced, it has had no appreciable impact on gorse invasiveness (Norambuena & Piper, 2000).

There are also examples that appear to show predator and prey populations linked together by coupled oscillations in abundance (Figure 10.1), but there are many more examples in which predator and prey populations fluctuate in abundance apparently independently of one another.

It is clearly a major task for ecologists to develop an understanding of the patterns of predator–prey abundance, and to account for the differences from one example to the next. It is equally clear, though, that none of these predator and prey populations exist as isolated pairs, but rather as parts of multispecies systems, and that all these species are affected by environmental conditions. These broader issues of what determines a species’ abundance are taken up again in Chapter 14. However, as with any complex process in science, we cannot understand the full complexity without a reasonable understanding of the components

– in this case, populations of predators and prey. Hence, this chapter deals with the consequences of predator–prey interactions for the dynamics of the populations concerned.

The approach will be firstly to use simple models to deduce the effects produced by different components of the interactions, teasing out the separate effects before seeking to understand those effects in combination. Then, field and experimental data will be examined to see whether the deductions appear to be supported or refuted. In fact, simple models are most useful when their predictions are *not* supported by real data – as long as the reason for the discrepancy can subsequently be discovered. Confirmation of a model’s predictions provides consolidation; refutation with subsequent explanation is progress.

10.2 The basic dynamics of predator–prey and plant–herbivore systems: a tendency towards cycles

There have been two main series of models developed as attempts to understand predator–prey dynamics. Both will be examined here. The first (Section 10.2.1) is based on differential equations (and hence, applies most readily to populations in which breeding is continuous), but relies heavily on simple graphical models (Rosenzweig & MacArthur, 1963). The second (Section 10.2.3) uses difference equations to model host–parasitoid interactions with discrete generations. Despite this taxonomic limitation, these models have the advantage of having been subject to rigorous mathematical exploration. (We have also noted previously that there are a very large number of important parasitoid species.) Although the two series of models are explained separately, they have, of course, a common aim (to advance our understanding of predator–prey dynamics), and they can increasingly be seen as ends of a discrete-to-continuous spectrum of mathematical approaches.

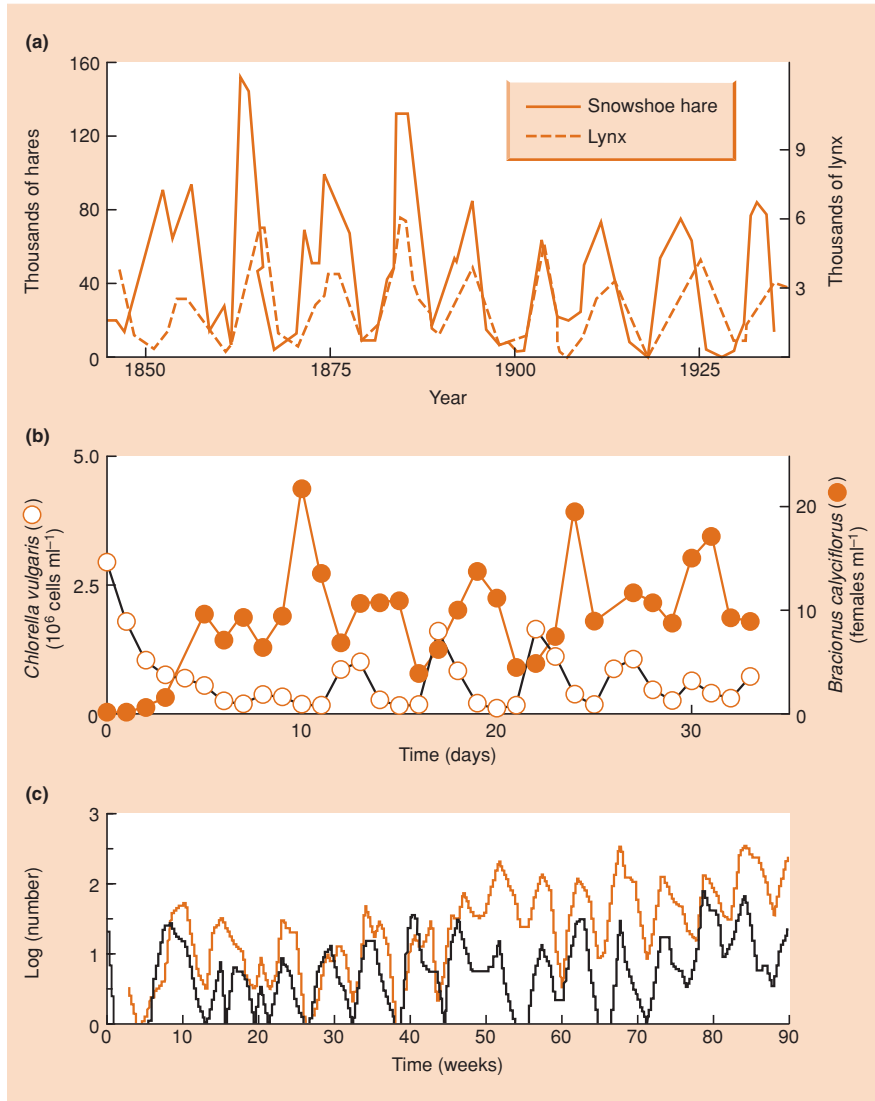


Figure 10.1 Coupled oscillations in the abundance of predators and prey. (a) The snowshoe hare (*Lepus americanus*) and the Canadian lynx (*Lynx canadensis*) as determined by the number of pelts lodged with the Hudson Bay Company. (After MacLulich, 1937.) (b) Parthenogenetic female rotifers, *Bracionus calyciflorus* (predators, ●), and unicellular green algae, *Chlorella vulgaris* (prey, ○) in laboratory cultures. (After Yoshida *et al.*, 2003.) (c) The parasitoid *Venturia canescens* (—) and its moth host *Plodia interpunctella* (—) in laboratory cultures. (After Bjørnstad *et al.*, 2001.)

10.2.1 The Lotka–Volterra model

The simplest differential equation model is known (like the model of interspecific competition) by the name of its originators: Lotka–Volterra (Volterra, 1926; Lotka, 1932). This will serve as a useful point of departure. The model has two components: P , the numbers present in a predator (or consumer) population, and N , the numbers or biomass present in a prey or plant population.

We assume initially that in the absence of consumers the prey population increases exponentially (see Section 5.9):

$$dN/dt = rN. \tag{10.1}$$

But prey individuals are removed by predators at a rate that depends on the frequency of predator–prey encounters. Encounters will

increase with the numbers of predators (P) and the numbers of prey (N). However, the exact number encountered and successfully consumed will depend on the searching and attacking efficiency of the predator: a , sometimes also called the ‘attack rate’. The consumption rate of prey will thus be aPN , and overall:

$$dN/dt = rN - aPN. \tag{10.2}$$

the Lotka–Volterra prey equation

In the absence of prey, predator numbers in the model are assumed to decline exponentially through starvation:

$$dP/dt = -qP, \tag{10.3}$$

where q is the predator mortality rate. This is counteracted by predator birth, the rate of which is assumed to depend on only

two things: the rate at which food is consumed, aPN , and the predator's efficiency, f , at turning this food into predator offspring. Predator birth rate is therefore $faPN$, and overall:

the Lotka–Volterra predator equation

$$dP/dt = faPN - qP. \quad (10.4)$$

Equations 10.2 and 10.4 constitute the Lotka–Volterra model.

The properties of this model can be investigated by finding zero isoclines. Zero isoclines were described for models of two-species competition in Section 8.4.1. Here, there are separate zero isoclines for the predators and prey, both of which are drawn on a graph of prey density (x -axis) against predator density (y -axis). Each is a line joining those combinations of predator and prey density that lead either to an unchanging prey population ($dN/dt = 0$; prey zero isocline) or an unchanging predator population ($dP/dt = 0$; predator zero isocline). Having drawn, say, a prey zero isocline, we know that combinations to one side of it lead to prey decrease, and combinations to the other to prey increase. Thus, as we shall see, if we plot the prey and predator zero isoclines on the same figure, we can begin to determine the pattern of the dynamics of the joint predator–prey populations.

In the case of the prey (Equation 10.2), when:

$$dN/dt = 0, rN = aPN \quad (10.5)$$

or:

$$P = r/a. \quad (10.6)$$

properties revealed by zero isoclines

Thus, since r and a are constants, the prey zero isocline is a line for which P itself is a constant (Figure 10.2a). Below it, predator abundance is low and the prey increase; above it, predator abundance is high and the prey decrease.

Likewise, for the predators (Equation 10.4), when:

$$dP/dt = 0, faPN = qP \quad (10.7)$$

or:

$$N = q/fa. \quad (10.8)$$

The predator zero isocline is therefore a line along which N is constant (Figure 10.2b). To the left, prey abundance is low and the predators decrease; to the right, prey abundance is high and the predators increase.

Putting the two isoclines together (Figure 10.2c) shows the behavior of joint populations. Predators increase in abundance when there are large numbers of prey, but this leads to an increased predation pressure on the prey, and thus to a decrease in prey abundance. This then leads to a food shortage for predators and a decrease in predator abundance, which leads to a relaxation of

predation pressure and an increase in prey abundance, which leads to an increase in predator abundance, and so on (Figure 10.2d). Thus, predator and prey populations undergo ‘coupled oscillations’ in abundance, which continue indefinitely.

The Lotka–Volterra model, then, is useful in pointing to this underlying tendency for predator–prey interactions to generate fluctuations in the prey population tracked by fluctuations in the predator population. The detailed behavior of the model, however, should not be taken seriously, because the cycles it exhibits are ‘structurally unstable’, showing ‘neutral stability’. That is, the populations would follow precisely the same cycles indefinitely, but only until some external influence shifted them to new values, after which they would follow new cycles indefinitely (Figure 10.2e). In practice, of course, environments are continually changing, and populations would continually be ‘shifted to new values’. A population following the Lotka–Volterra model would, therefore, not exhibit regular cycles, but, because of repeated disturbance, fluctuate erratically. No sooner would it start one cycle than it would be diverted to a new one.

an underlying tendency towards coupled oscillations – which are structurally unstable in this case

For a population to exhibit regular and recognizable cycles, the cycles must themselves be stable: when an external influence changes the population level, there must be a tendency to return to the original cycle. In fact, as we shall see, predator–prey models (once we move beyond the very limiting assumptions of the Lotka–Volterra model) are capable of generating a whole range of abundance patterns: stable-point equilibria, multigeneration cycles, one-generation cycles, chaos, etc. – a range repeated in surveys of real populations. The challenge is to discover what light the models can throw on the behavior of real populations.

10.2.2 Delayed density dependence

The basic mechanism generating the coupled oscillations in these predator–prey interactions is a series of time-delayed ‘numerical responses’, i.e. changes in one species’ abundance in response to the abundance of the other species. The first is a time delay between ‘many prey’ and ‘many predators’ (arising because the response of predator abundance to high prey abundance cannot occur instantaneously). There may then be another time delay between ‘many predators’ and ‘few prey’, and then between ‘few prey’ and ‘few predators’, and so on. In practice, therefore, even where coupled oscillations exist, their exact shape is likely to reflect the varying delays, and strengths, of the different numerical responses. Certainly, the shapes of apparent coupled oscillations in real populations are varied, and not all are symmetric like those generated by the Lotka–Volterra model (see Figure 10.1).

numerical responses

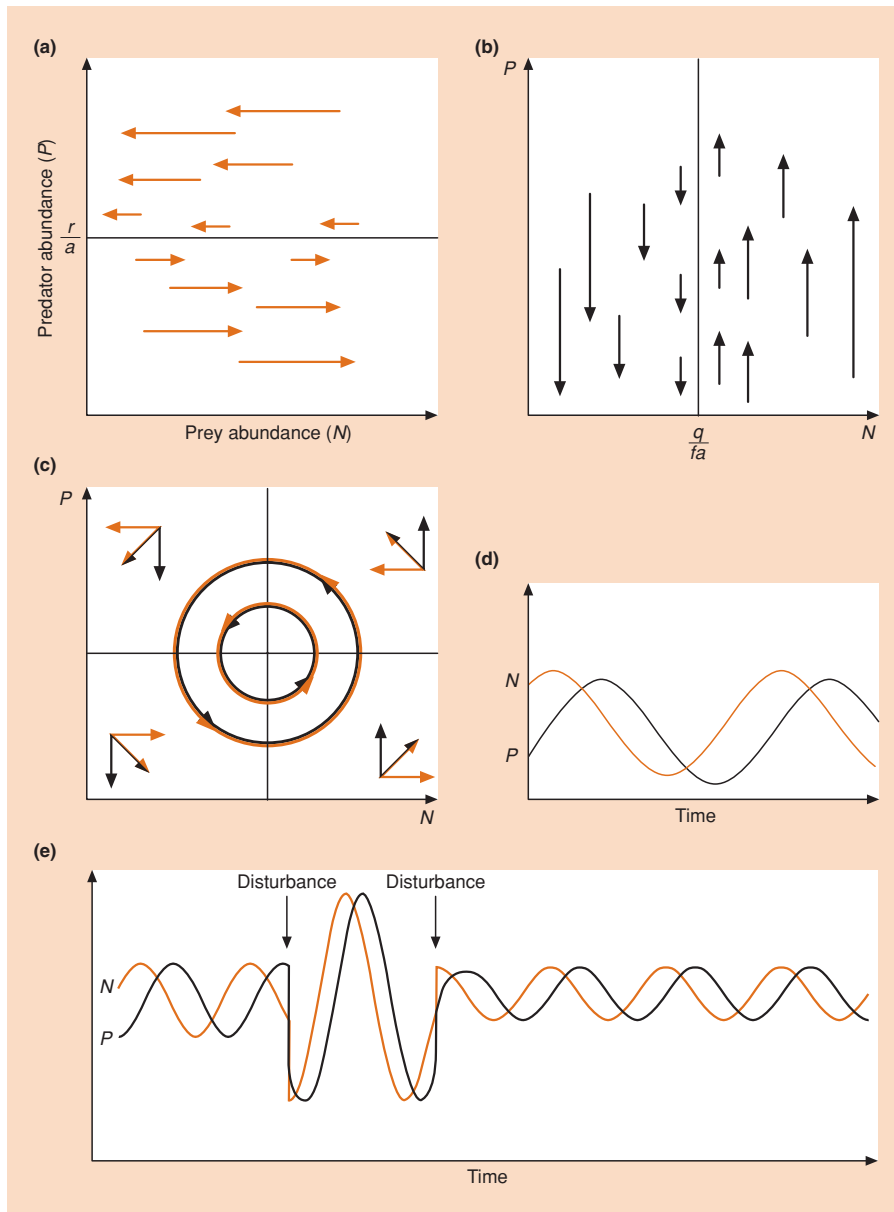


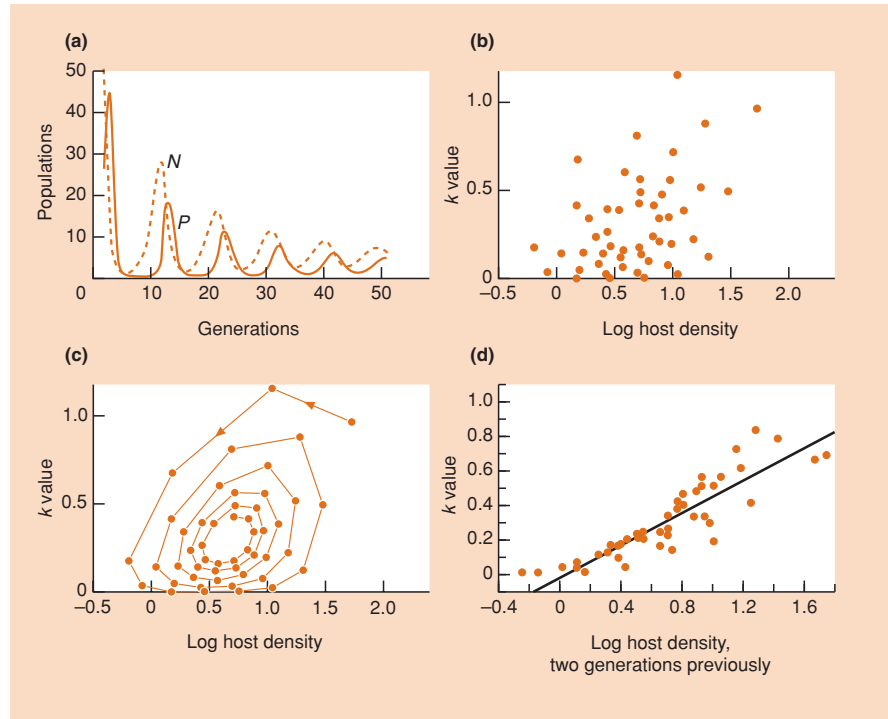
Figure 10.2 The Lotka–Volterra predator–prey model. (a) The prey zero isocline, with prey (N) increasing in abundance (arrows left to right) at lower predator densities (low P) and decreasing at higher predator densities. (b) The predator zero isocline, with predators increasing in abundance (arrows pointing upwards) at higher prey densities and decreasing at lower prey densities. (c) When the zero isoclines are combined, the arrows can also be combined, and these joint arrows progress in anticlockwise circles. In other words, the joint population moves with time from low predator/low prey (bottom left in (c)), to low predator/high prey (bottom right), to high predator/high prey, to high predator/low prey and back to low predator/low prey. Note, however, that the lowest prey abundance (‘9 o’clock’) comes one-quarter of a cycle before the lowest predator abundance (‘6 o’clock’ – anticlockwise movement). These coupled cycles of predator–prey abundance, continuing indefinitely, are shown as numbers against time in (d). However, as shown in (e), these cycles exhibit neutral stability: they continue indefinitely if undisturbed, but each disturbance to a new abundance initiates a new, different series of neutrally stable cycles, around the same means but with a different amplitude.

the regulatory tendencies of delayed density dependence are relatively difficult to demonstrate

These responses are density dependent (see Section 5.2): they act to reduce the size of relatively large populations and allow relatively small populations to increase. Varley (1947) introduced the term ‘delayed density dependence’ to describe them. The strength of a delayed density-dependent effect is related not to the current abundance (that would be *direct* density dependence) but to abundance at some time in the past (i.e. the delay-length ago). Compared to direct density

dependence, delayed density dependence is relatively difficult to demonstrate. To see this, we can examine the coupled oscillations produced by a particular parasitoid–host model, shown in Figure 10.3a (Hassell, 1985). The details of the model need not concern us, but note that the oscillations are damped: they get gradually smaller over time until a stable equilibrium is reached. The prey population, subject to delayed density dependence, is regulated in size by the predator. In Section 5.6, we demonstrated density dependence by plotting k values against the log of density; but when we plot the k values of predator-induced mortality

Figure 10.3 Delayed density dependence. (a) A parasitoid–host model followed over 50 generations: despite oscillations, the parasitoid has a regulatory effect on the host population. (b) For the same model, the k value of generation mortality plotted against the log of host density: no clear density-dependent relationship is apparent. (c) The points from (b) linked serially from generation to generation: they spiral in an anticlockwise direction – a characteristic of delayed density dependence. (After Hassell, 1985). (d) The k value of generation mortality plotted against the log of host density two generations previously: a clear delayed density-dependent relationship is again apparent.



against the log of prey density in that generation (Figure 10.3b), no clear relationship is apparent. On the other hand, when the same points are linked together, each generation to the next (Figure 10.3c), they can be seen to describe an anticlockwise spiral. This spiraling is characteristic of delayed density dependence. Here, because the oscillations are damped, the points spiral inwards to the equilibrium point. Moreover, when we plot the k values of predator-induced mortality against the log of prey density *two generations previously* (Figure 10.3d), the delayed density dependence is clearly revealed by the positive relationship characteristic of density dependence in general. Indeed, the fact that a two-generation delay gives a better fitting relationship than delays that are either shorter or longer, tells us that two generations is our best estimate of the delay in this case.

The regulatory effects of delayed density dependence are relatively easy to reveal for the model population of Figure 10.3, because it is not subject to the fluctuations of a natural environment, it is not subject to the density-dependent attacks of any other predator, it is not subject to the inaccuracies of sampling error, and so on. Data of this quality, however, are rarely if ever available for natural or even experimental populations. We return to the question of uncovering and integrating delayed density-dependent effects into an overall account of what determines abundance in Chapter 14. For now, though, this discussion

highlights the relationship between ‘regulation’ and ‘stability’ in predator–prey interactions. Natural predator and prey populations tend to exhibit less violent and less regular fluctuations than those we have seen generated by the simplest models. Most of the rest of this chapter describes the search for explanations for these patterns and for the variations in dynamical pattern from case to case. A population that remains roughly constant in size provides evidence for the effects of both regulatory and stabilizing forces. The delayed density dependence of a predator–prey interaction ‘regulates’ in the sense of acting strongly on large populations and only weakly on small populations. But, as we have already seen, it can hardly be said, typically, to stabilize either population. What follows in this chapter is, therefore, in large part, a search for stabilizing forces that might complement the (delayed) regulatory forces that occur inherently in predator–prey interactions.

10.2.3 The Nicholson–Bailey model

Turning now to parasitoids, the basic model (Nicholson & Bailey, 1935) is again not so much realistic as a reasonable basis from which to start. Let H_t be the number of hosts, and P_t the number of parasitoids in generation t ; r is the intrinsic rate of natural increase of the host. If H_a is the number of hosts attacked by

parasitoids (in generation t), then, assuming no intraspecific competition amongst the hosts (exponential growth – see Section 4.7.1), and that each host can support only one parasitoid (commonly the case):

$$H_{t+1} = e^r(H_t - H_a), \quad (10.9)$$

$$P_{t+1} = H_a. \quad (10.10)$$

In other words, hosts that are not attacked reproduce, and those that are attacked yield not hosts but parasitoids.

To derive a simple formulation for H_a , let E_t be the number of host–parasitoid encounters in generation t . Then, if A is the parasitoid’s searching efficiency:

$$E_t = AH_tP_t \quad (10.11)$$

and:

$$E_t/H_t = AP_t. \quad (10.12)$$

Note the similarity to the formulation in Equation 10.2. Remember, though, that we are dealing with parasitoids, and hence a single host can be encountered several times, although only one encounter leads to successful parasitization (i.e. only one parasitoid develops). Predators, by contrast, would remove their prey and prevent re-encounters. Thus, Equation 10.2 dealt with instantaneous rates, rather than numbers.

a model based on random encounters . . .

If encounters are assumed to occur more or less at random, then the proportions of hosts that are encountered zero, one, two or more times are given by the successive terms in the appropriate ‘Poisson distribution’ (see any basic statistics textbook). The proportion not encountered at all, p_0 , would be given by e^{-E_t/H_t} , and thus the proportion that is encountered (one or more times) is $1 - e^{-E_t/H_t}$. The number encountered (or attacked) is then:

$$H_a = H_t(1 - e^{-E_t/H_t}). \quad (10.13)$$

Using this and Equation 10.12 to substitute into Equations 10.9 and 10.10 gives us:

$$H_{t+1} = H_t e^{r(1 - AP_t)} \quad (10.14)$$

$$P_{t+1} = H_t(1 - e^{-AP_t}). \quad (10.15)$$

. . . giving rise to (unstable) coupled oscillations

This is the basic Nicholson–Bailey model of a host–parasitoid interaction. Its behavior is reminiscent of the Lotka–Volterra model but it is even less stable. An equilibrium combination

of the two populations is a possibility, but even the slightest disturbance from this equilibrium leads to divergent coupled oscillations.

10.2.4 One-generation cycles

The coupled oscillations generated by the basic Lotka–Volterra and Nicholson–Bailey models are multigeneration cycles, i.e. there are several generations between successive peaks (or troughs), and such oscillations have lain at the heart of most attempts to understand cyclic predator–prey dynamics. However, other models of host–parasitoid (and host–pathogen) systems are able to generate coupled oscillations just *one* host generation in length (Knell, 1998; see, for example, Figure 10.1c). On the other hand, such ‘generation cycles’ can also occur in a population for reasons other than a predator–prey interaction – specifically as a result of competition between age classes within a population (Knell, 1998).

Predator–prey generation cycles occur essentially when the generation length of the consumer is roughly half that of its host – as it often is. Any small, chance peak in host abundance tends to generate a further peak in host abundance one host generation later. But any associated peak in consumer abundance occurs half a host generation length later, creating a trough in host abundance between the twin peaks. And this host trough creates a further host trough one generation later, but a consumer trough coinciding with the next host peak. Thus, the consumers have alternate ‘feasts’ and ‘famines’ that accentuate the originally small peaks and troughs in host abundance, and hence promote one-generation cycles (Figure 10.4).

10.2.5 Predator–prey cycles in nature: or are they?

The inherent tendency for predator–prey interactions to generate coupled oscillations in abundance might suggest an expectation of such oscillations in real populations. However, there are many important aspects of predator and prey ecology that have not been considered in the models derived so far; and as subsequent sections will show, these can greatly modify our expectations. Certainly, even if a population exhibits regular oscillations, this does not necessarily provide support for the Lotka–Volterra, Nicholson–Bailey or any other simple model. We saw cycles generated by intraspecific competition in Section 5.8, and we shall see several other routes to cycles in subsequent chapters (see also Kendall *et al.*, 1999). At this point, though, it is worth simply making the point that even when predators or prey exhibit regular cycles in abundance, it is never easy to demonstrate that these are *predator–prey* cycles.

hare and lynx: not the simple predator and prey they appear to be

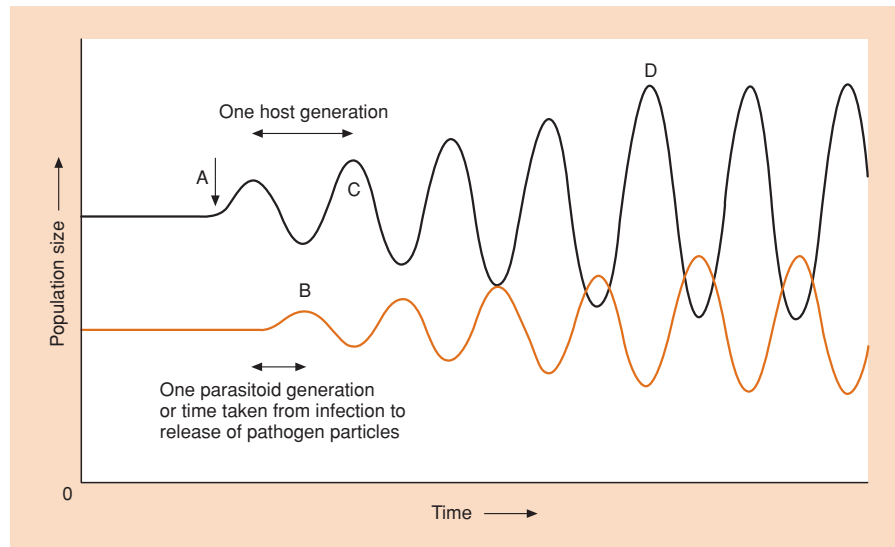


Figure 10.4 Schematic illustration of how a parasitoid or pathogen may generate coupled cycles in abundance in the host and itself that are approximately one host generation in length. For this, the parasitoid or pathogen must have a generation length approximately half that of the host. Any chance increase in host abundance (A) will first give rise to an increase in parasitoid abundance one parasitoid generation later (B), and also to an increase in host abundance one host generation later (C). But the parasitoid peak at B will also give rise to a coincident host trough, which will give rise to a parasitoid trough at C, reinforcing the host peak at that point. This mutual reinforcement will continue until by, say, D, persistent host generation-length cycles have become established. (After Knell, 1998; from Godfray & Hassell, 1989.)

The regular oscillations in the abundance of the snowshoe hare and the Canadian lynx shown in Figure 10.1a have often been held to epitomize predator–prey cycles. Recently, however, evidence has increasingly indicated that even this apparent exemplar is not as straightforward as it has seemed. Experimental manipulations carried out in the field are one powerful means of suggesting what forces are normally acting: if those forces are removed or exaggerated, is the cycle eliminated or enhanced? A whole series of coordinated field experiments has indicated that the cyclic hare is not simply a prey of the lynx (and other predators in the community), nor simply a predator of its plant food resources: the cycle can be understood only by taking account of its interactions both as a prey *and* as a predator (Krebs *et al.*, 2001). Furthermore, modern statistical analysis of the time series of abundances has tended to confirm this: the hare series carries a relatively complex ‘signature’, suggesting the influence of both its predators and its food, whereas the lynx series has a simpler signature, suggesting only the influence of its (hare) prey (Stenseth *et al.*, 1997; see also Section 14.5.2). What has so often been described as a predator–prey cycle seems rather to comprise one predator linked to a species that is both predator and prey.

moths and two natural enemies

Apparently coupled one-generation cycles linking a moth host (*Plodia interpunctella*) and its parasitoid *Venturia canescens* were shown in Figure 10.1c.

In this case, the dangers of jumping too readily to the conclusion

that these are predator–prey cycles are highlighted by the fact that the host also exhibits generation-length cycles when maintained alone, without any natural enemies, and also when maintained with another enemy, a granulovirus (Figure 10.5). It has been possible, however, to confirm that the cycles in Figure 10.1c are indeed coupled oscillations, using methods similar to those applied to the hare–lynx time series (Bjørnstad *et al.*, 2001). The host-alone cycles have within them the signature simply of intraspecific competition, and the virus seems to modulate this pattern but does not alter its basic structure (i.e. the patterns in Figure 10.5 are *not* predator–prey cycles). However, the host and parasitoid cycles in Figure 10.1c both carry the same, more complex signature that indicates a tightly coupled prey–predator interaction (see also Section 12.7.1).

We return to the question of cycles – indeed, some of the same cycles discussed above – in Section 14.6, as part of a more general exploration of how the whole range of biotic and abiotic factors come together to determine the level and pattern of a population’s abundance.

10.3 Effects of crowding

The most obvious omission, perhaps, from the predator–prey interactions we have modeled so far has been any acknowledgement that prey abundance may be limited by other prey, and predator

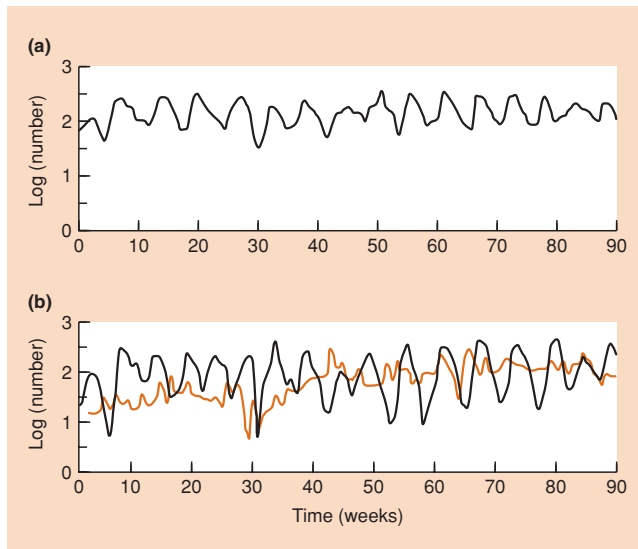


Figure 10.5 Host generation-length cycles in the moth *Plodia interpunctella* (a) alone (black line) and (b) with a granulovirus (colored line). These dynamics may be compared with those in Figure 10.1c. In spite of a superficial similarity in pattern, analysis indicates that those in (a) are generated by intraspecific competition; those in (b) are simply modulated versions of those in (a) and are therefore not predator–prey cycles. However, those in Figure 10.1c are predator–prey cycles. (After Bjornstad *et al.*, 2001.)

abundance by other predators. Prey are bound to be increasingly affected by intraspecific competition as their abundance increases; and predators, too, are likely to be limited at high densities by the availability of resting places, say, or safe refuges of their own, quite apart from their interaction with their most obvious resource, their prey.

mutual interference

More generally, predators have been assumed in the models discussed thus far to consume prey at a rate that depends only on prey abundance (in Equation 10.2, for example, the consumption rate per predator is simply aN). In reality, consumption rate will also often depend on the abundance of the predators themselves. Most obviously, food shortage – the abundance of prey *per predator* – will commonly result in a reduction in the consumption rate per individual as predator density increases. However, even when food is not limited, the consumption rate can be reduced by a number of processes known collectively as mutual interference (Hassell, 1978). For example, many consumers interact behaviorally with other members of their population, leaving less time for feeding and therefore depressing the overall feeding rate. For instance, humming-birds actively and aggressively defend rich sources of nectar. Alternatively, an increase in consumer density may lead to an increased rate of emigration, or of

consumers stealing food from one another (as do many gulls), or the prey themselves may respond to the presence of consumers and become less available for capture. All of these mechanisms give rise to a decline in predator consumption rate with predator density. Figure 10.6a, for example, shows significant reductions in consumption rate with abundance even at low densities of the crab *Carcinus aestuarii* foraging for the mussel *Musculista senhousia*; while Figure 10.6b shows that the kill rate of wolves, *Canis lupus*, preying on moose, *Alces alces*, in Isle Royale National Park, Michigan, USA, was lowest when there were most wolves.

10.3.1 Crowding in the Lotka–Volterra model

The effects of intraspecific competition, and of a decline in predator consumption rate with predator density, can be investigated by modifying the Lotka–Volterra isoclines. The details of incorporating intraspecific competition into the prey zero isocline are described by Begon *et al.* (1990), but the end result (Figure 10.7a) can be understood without reference to these details. At low prey densities there is no intraspecific competition, and the prey isocline is horizontal as in the Lotka–Volterra model. But as density increases, it is increasingly the case that prey densities below the isocline (prey increase) must be placed above the isocline (prey decrease) because of the effects of intraspecific competition. Hence, the isocline is increasingly lowered until it reaches the prey axis at the carrying capacity, K_N ; that is, the prey can only just maintain themselves even in the absence of predators.

As we have seen, the predator isocline in the Lotka–Volterra model is vertical. This itself reflects the assumption that the ability of a predator population to increase in abundance is determined

crowding and the Lotka–Volterra isoclines

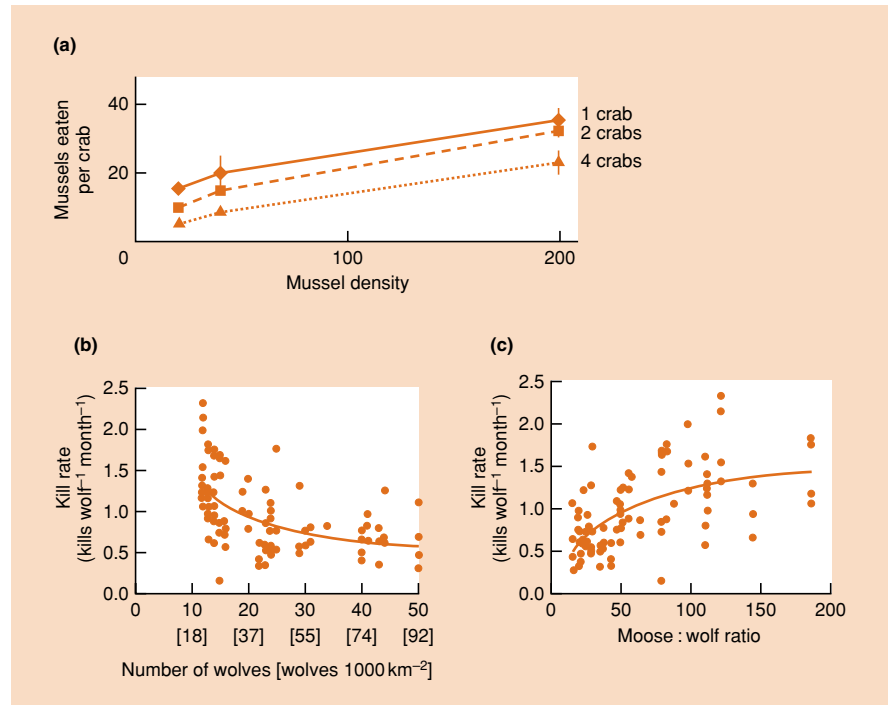
by the absolute abundance of prey, irrespective of the number of predators. If, however, mutual interference amongst the predators increases, then individual consumption rates will decline with predator abundance, and additional prey will be required to maintain a predator population of any given size. The predator zero isocline will depart increasingly from the vertical (Figure 10.7b). Moreover, at high densities, competition for other resources will put an upper limit on the predator population (a horizontal isocline) irrespective of prey numbers (Figure 10.7b).

An alternative modification is to abandon altogether the assumption that consumption rate depends only on the absolute availability of prey, and assume

ratio-dependent predation

ratio-dependent predation instead (Arditi & Ginzburg, 1989), although this alternative has itself been criticized (see Abrams, 1997; Vucetich *et al.*, 2002). In this case, the consumption rate depends on the ratio of prey to predators, and a particular ratio needs to be exceeded for the predators to increase in abundance: a

Figure 10.6 (a) Mutual interference amongst crabs, *Carcinus aestuarii*, feeding on mussels, *Musculista senhousia*. ♦ 1 crab; ■, 2 crabs; ▲, 4 crabs. The more crabs there were, the lower their per capita consumption rate. (After Mistri, 2003.) (b) Mutual interference amongst wolves, *Canis lupus*, preying on moose, *Alces alces*. (c) The same data but with wolf kill rate plotted against the moose : wolf ratio. The fitted curve assumes a dependence of kill rate on this ratio, but also that the wolves may become 'saturated' at high moose densities (see Section 10.4.2). This curve fits better than any for which kill rate depends on either predator density (e.g. (b)) or prey density. ((b, c) after Vucetich *et al.*, 2002.)



diagonal zero isocline passing through the origin (Figure 10.7c). Evidence of ratio-dependent predation is illustrated, for example, for the wolf–moose study in Figure 10.6c.

The likely effects of crowding in either population can now be deduced by combining the predator and prey isoclines (Figure 10.7d). Oscillations are still apparent for the most part, but these are no longer neutrally stable. Instead, they are damped so that they converge to a stable equilibrium. Predator–prey interactions in which either or both populations are substantially self-limited are likely, therefore, to exhibit patterns of abundance that are relatively stable, i.e. in which fluctuations in abundance are relatively slight.

crowding stabilizes dynamics

More particularly, when the predator is relatively inefficient, i.e. when many prey are needed to maintain a population of predators (curve (ii) in Figure 10.7d), the oscillations are damped quickly but the equilibrium prey abundance (N^*) is not much less than the equilibrium in the absence of predators (K_N). By contrast, when the predators are more efficient (curve (i)), N^* is lower and the equilibrium density of predators, P^* , is higher – but the interaction is less stable (the oscillations are more persistent). Moreover, if the predators are very strongly self-limited, then abundance may not oscillate at all (curve (iii)); but P^* will tend to be low, whilst N^* will tend to be not much less than K_N . Hence, for interactions where there is crowding, there appears to be a contrast between those in which predator density is low, prey abundance is little affected and the patterns of abundance are stable, and those in which predator

density is higher and prey abundance is more drastically reduced, but the patterns of abundance are less stable. (Figure 10.7d does not use ratio-dependent predation, but a predator isocline with a steeper slope in a ratio-dependent model (more efficient predation) can be equated, for present purposes, with an isocline rising from closer to the origin in the figure – that is, curve (i) rather than curve (ii).)

Essentially similar conclusions emerge from modifications of the Nicholson–Bailey model that incorporate either simple (logistic) crowding effects amongst the hosts or mutual interference amongst the predators (Hassell, 1978).

To quote examples of data proving the stabilizing influence of self-limitation on predator–prey dynamics would be difficult, simply because it would be all but impossible to compare the dynamics of matched populations with and without such self-limitation. On the other hand, populations of predators and prey with relatively stable dynamics are commonplace, as are the stabilizing forces of self-limitation we have discussed here. To take a more specific example, there are two groups of primarily herbivorous rodents that are widespread in the Arctic: the microtine rodents (lemmings and voles) and the ground squirrels. The microtines are renowned for their dramatic, cyclic fluctuations in abundance (see Chapter 14), but the ground squirrels have populations that remain remarkably constant from year to year, especially in open meadow and tundra habitats. There, significantly, they appear to be strongly self-limited by food availability, suitable burrowing habitat and their own spacing behavior (Karels & Boonstra, 2000).

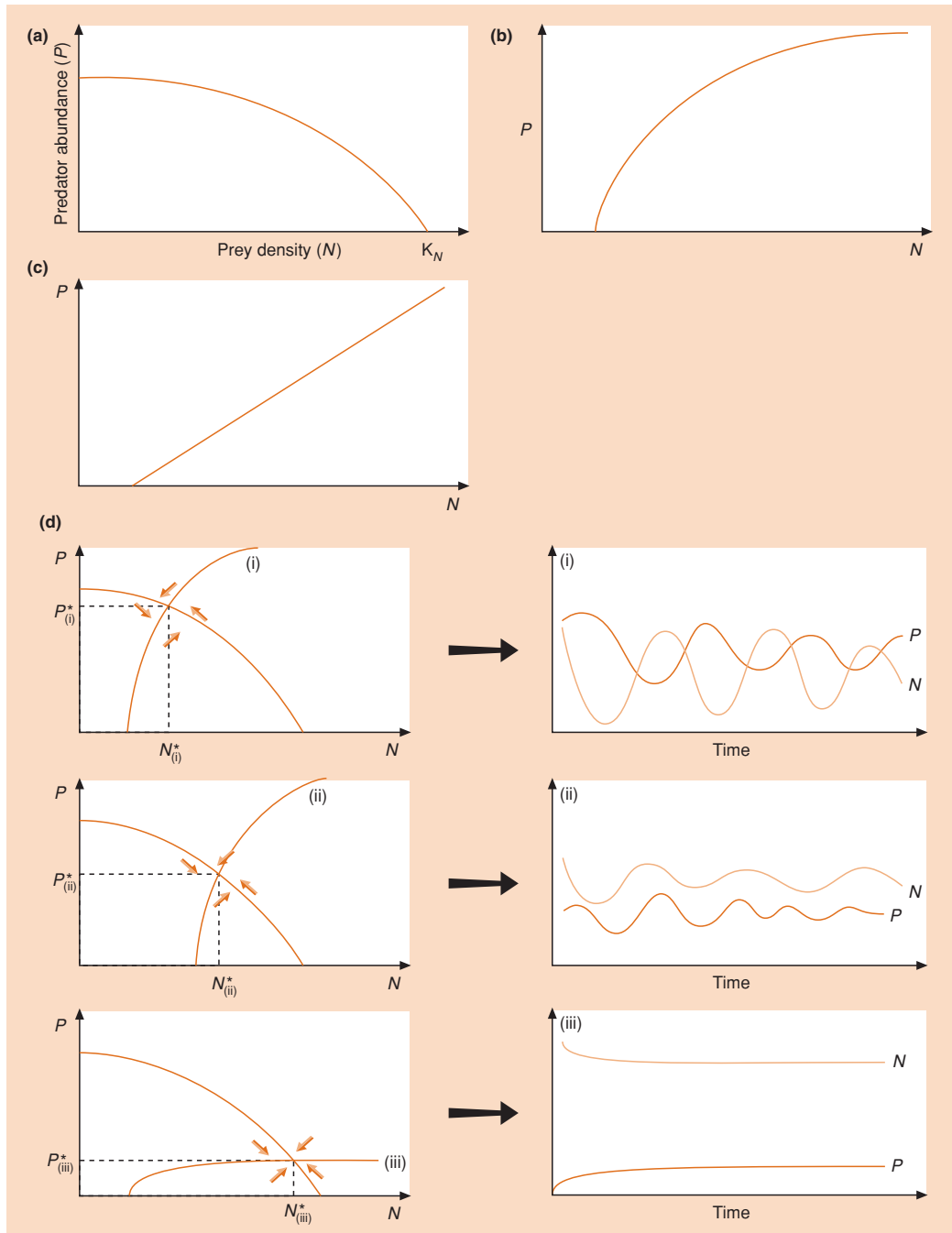


Figure 10.7 (a) A prey zero isocline subject to crowding. At the lowest prey densities this is the same as the Lotka–Volterra isocline, but when the density reaches the carrying capacity (K_N) the population can only just maintain itself even in the complete absence of predators. (b) A predator zero isocline subject to crowding (see text). (c) A predator zero isocline when there is prey : predator ratio dependent predation. (d) The prey zero isocline combined with the predator zero isoclines with increasing levels of crowding: (i), (ii) and (iii). P^* is the equilibrium abundance of predators, and N^* the equilibrium abundance of prey. Combination (i) is the least stable (most persistent oscillations) and has the most predators and least prey: the predators are relatively efficient. Less efficient predators, as in (ii), give rise to a lowered predator abundance, an increased prey abundance and less persistent oscillations. Strong predator self-limitation (iii) can eliminate oscillations altogether, but P^* is low and N^* is close to K_N .

how important is mutual interference in practice?

On a cautionary note, however, Umbanhowar *et al.* (2003), for example, failed to find evidence of mutual interference in a field study of the parasitoid *Tachinomyia similis* attacking its moth host *Orgyia vetusta*. The strength of mutual interference may often have been exaggerated by forcing predators to forage in artificial arenas at densities much higher than those they experience naturally. This is a useful reminder of the general point that an ecological force that is powerful in models or in the laboratory may none the less often be trivial, in practice, in natural populations. There can be little doubt, though, that self-limitation in its various forms frequently plays a key role in shaping predator-prey dynamics.

10.4 Functional responses

Having examined the relationship between a predator's consumption rate and the predator's own abundance, above, we turn now to the effect on this consumption rate of the prey's abundance, the so-called functional response (Solomon, 1949). Below we describe the three main types of functional response (Holling, 1959), before considering how they might modify predator-prey dynamics.

10.4.1 The type 1 functional response

The most basic, 'type 1' functional response is that assumed by the Lotka–Volterra equations: consumption rate rises linearly with prey density (indicated by the constant, a , in Equation 10.2). An example is illustrated in Figure 10.8. The rate at which *Daphnia magna* consumed yeast cells rose linearly when the density of cells

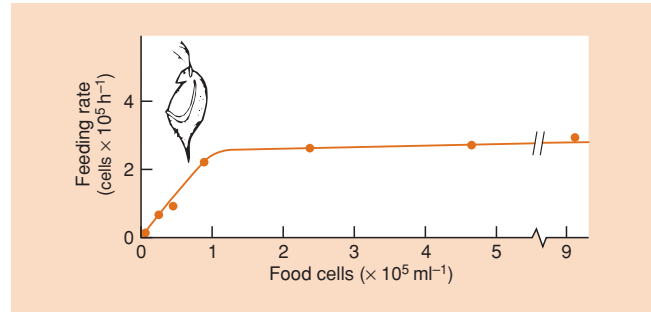


Figure 10.8 The type 1 functional response of *Daphnia magna* to different concentrations of the yeast *Saccharomyces cerevisiae*. (After Rigler, 1961.)

varied, because the yeast cells were extracted by the *Daphnia* from a constant volume of water washed over their filtering apparatus, and the amount extracted therefore rose in line with food concentration. Above 10^5 cells ml^{-1} , however, the *Daphnia* could filter more cells but were unable to swallow all the food they filtered. They therefore ingested food at a maximum (plateau) rate irrespective of its concentration.

10.4.2 The type 2 functional response

The most frequently observed functional response is the 'type 2' response, in which consumption rate rises with prey density, but gradually decelerates until a plateau is reached at which consumption rate remains constant irrespective of prey density. (Realistically, even a type 1 response must have a plateau, as in the example above. The distinction is between the deceleration of a type 2 response and the linearity of the type 1 response.) Type 2 responses are shown for a carnivore, a herbivore and a parasitoid in Figure 10.9.

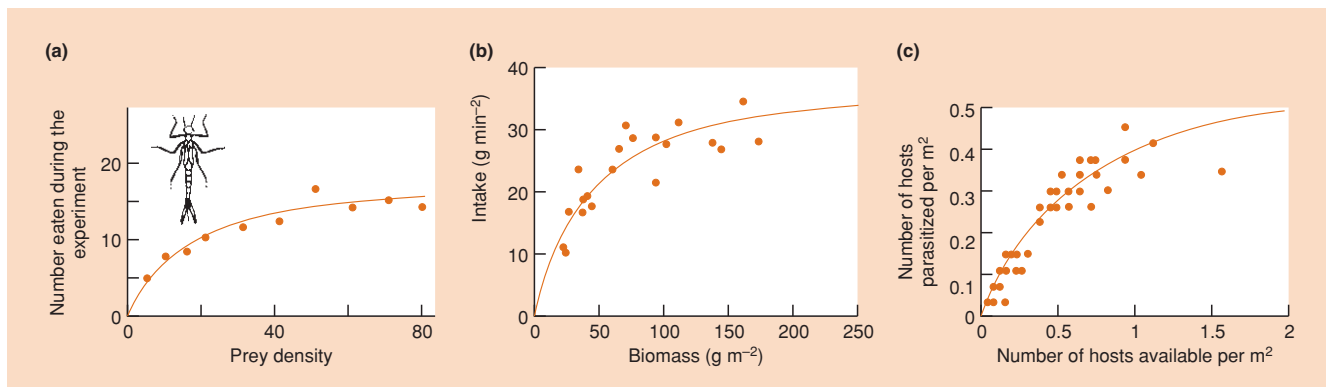


Figure 10.9 Type 2 functional responses. (a) Tenth-instar damselfly nymphs (*Ishnura elegans*) eating *Daphnia* of approximately constant size. (After Thompson, 1975.) (b) Wood bison (*Bison bison*) feeding on the sedge *Carex atherodes* presented at a range of sedge biomass densities. (After Bergman *et al.*, 2000.) (c) The parasitoid *Microplitis croceipes* attacking the tobacco budworm *Heliothis virescens*. (After Tillman, 1996.)

the type 2 response and handling time

The type 2 response can be explained by noting that a predator has to devote a certain handling time to each prey item it consumes (i.e. pursuing, subduing and consuming the prey item, and then preparing itself for further search). As prey density increases, finding prey becomes increasingly easy. Handling a prey item, however, still takes the same length of time, and handling overall therefore takes up an increasing proportion of the predator's time – until at high prey densities the predator is effectively spending all of its time handling prey. The consumption rate therefore approaches and then reaches a maximum (the plateau), determined by the maximum number of handling times that can be fitted into the total time available.

We can derive a relationship between P_c (the number of prey items eaten by a predator during a period of searching time, T_s) and N , the density of those prey items (Holling, 1959). P_c increases with the time available for searching, it increases with prey density, and it increases with the searching efficiency or attack rate of the predator, a . Thus:

Holling's type 2 response equation

$$P_c = aT_s N. \quad (10.16)$$

However, the time available for searching will be less than the total time, T , because of time spent handling prey. Hence, if T_h is the handling time of each prey item, then $T_h P_c$ is the total time spent handling prey, and:

$$T_s = T - T_h P_c. \quad (10.17)$$

Substituting this into Equation 10.16 we have:

$$P_c = a(T - T_h P_c)N \quad (10.18)$$

or, rearranging:

$$P_c = aNT / (1 + aT_h N). \quad (10.19)$$

Note that the equation describes the amount eaten during a specified period of time, T , and that the density of prey, N , is assumed to remain constant throughout that period. In experiments, this can sometimes be guaranteed by replacing any prey that are eaten, but more sophisticated models are required if prey density is depleted by the predator. Such models are described by Hassell (1978), who also discusses methods of estimating attack rates and handling times from a set of data. (Trexler *et al.*, 1988, discuss the general problem of fitting functional response curves to sets of data.)

other routes to a type 2 response

It would be wrong, however, to imagine that the existence of a handling time is the only or the complete explanation for all type 2 functional

responses. For instance, if the prey are of variable profitability, then at high densities the diet may tend towards a decelerating number of highly profitable items (Krebs *et al.*, 1983); or a predator may become confused and less efficient at high prey densities.

10.4.3 The type 3 functional response

Type 3 functional responses are illustrated in Figure 10.10a–c. At high prey densities they are similar to a type 2 response, and the explanations for the two are the same. At low prey densities, however, the type 3 response has an accelerating phase where an increase in density leads to a more than linear increase in consumption rate. Overall, therefore, a type 3 response is 'S-shaped' or 'sigmoidal'.

switching

One important way in which a type 3 response can be generated is through switching by the predator (see Section 9.5.2). The similarities between Figures 9.15 and 10.10 are readily apparent. The difference is that discussions of switching focus on the density of a prey type relative to the densities of alternatives, whereas functional responses are based on only the absolute density of a single prey type. In practice, though, absolute and relative densities are likely to be closely correlated, and switching is therefore likely to lead frequently to a type 3 functional response.

variations in searching efficiency or handling time

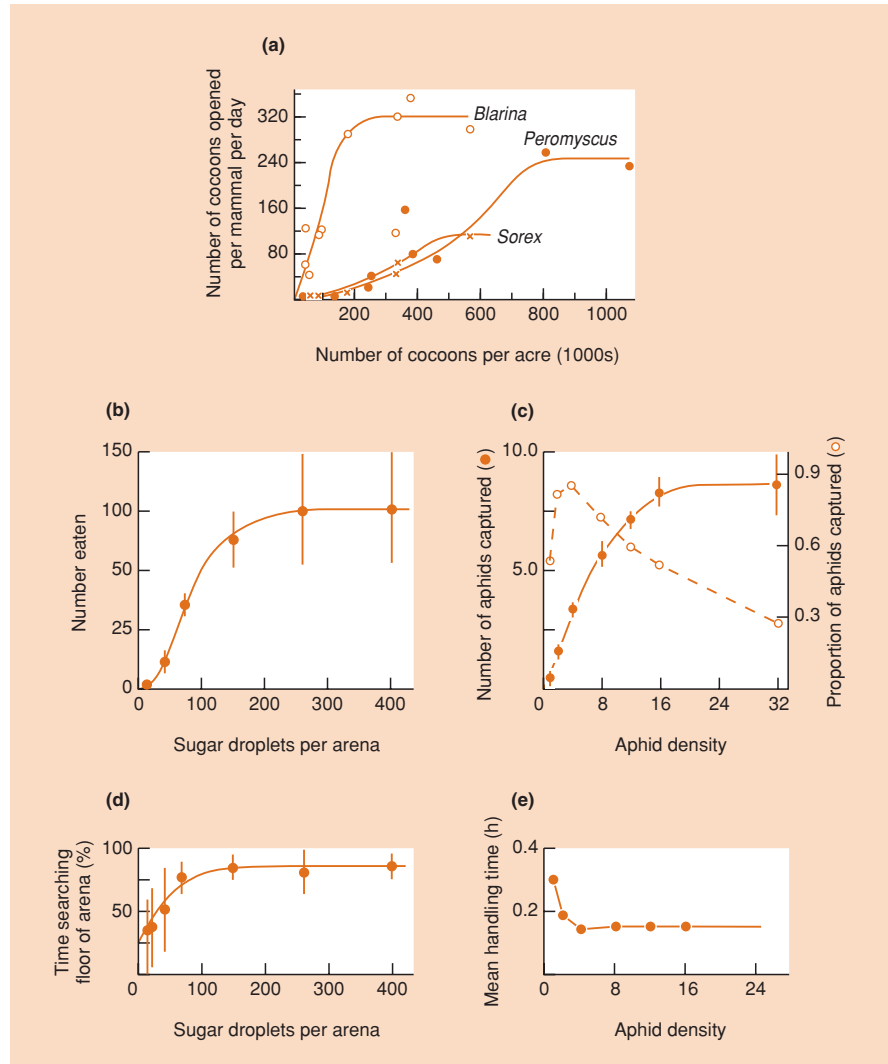
More generally, a type 3 functional response will arise whenever an increase in food density leads to an increase in the consumer's searching efficiency, a , or a decrease in its handling time, T_h , since between them these two determine consumption rate (Equation 10.19). Thus, the small mammals in Figure 10.10a appear to develop a search image for sawfly cocoons as they become more abundant (increasing efficiency). The bluebottle fly, *Calliphora vomitoria* (Figure 10.10b), spends an increasing proportion of its time searching for 'prey' as prey density increases (Figure 10.10d), also increasing efficiency. Whilst the wasp *Aphelinus thomsoni* (Figure 10.10c) exhibits a reduction in mean handling time as the density of its sycamore aphid prey increases (Figure 10.10e). In each case, a type 3 functional response is the result.

10.4.4 Consequences for population dynamics of functional responses and the Allee effect

Different types of functional response have different effects on population dynamics. A type 3 response means a low predation rate at low prey densities. In terms of isoclines, this means that

type 3 responses stabilize but may be unimportant in practice

Figure 10.10 Type 3 (sigmoidal) functional responses. (a) The shrews *Sorex* and *Blarina* and the deer mouse *Peromyscus* responding to changing field densities of cocoons of the European pine sawfly, *Neodiprion sertifer*, in Ontario, Canada. (After Holling, 1959.) (b) The bluebottle fly, *Calliphora vomitoria*, feeding on sugar droplets. (After Murdie & Hassell, 1973.) (c) The wasp, *Aphelinus thomsoni*, attacking sycamore aphids, *Drepanosiphum platanoidis*: note the density-dependent increase in prey mortality rate at low prey densities (---) giving rise to the accelerating phase of the response curve (—). (After Collins *et al.*, 1981.) (d) The basis of the response in (b): searching efficiency of *C. vomitoria* increases with ‘prey’ (sugar droplet) density. (After Murdie & Hassell, 1973.) (e) The basis of the response in (c): handling time in *A. thomsoni* decreases with aphid density. (After Collins *et al.*, 1981.)



prey at low densities can increase in abundance virtually irrespective of predator density, and that the prey zero isocline will therefore rise vertically at low prey densities (Figure 10.11a). This could lend considerable stability to an interaction (Figure 10.11a, curve (i)), but for this the predator would have to be highly efficient at low prey densities (readily capable of maintaining itself), which contradicts the whole idea of a type 3 response (ignoring prey at low densities). Hence, curve (ii) in Figure 10.11a is likely to apply, and the stabilizing influence of the type 3 response may in practice be of little importance.

On the other hand, if a predator has a type 3 response to one particular type of prey because it switches its attacks amongst various prey types, then the population dynamics of the predator would be independent of the abundance of any particular prey type, and the vertical position of its zero isocline would therefore

be the same at all prey densities. As Figure 10.11b shows, this can lead potentially to the predators regulating the prey at a low and stable level of abundance.

An apparent example of this is provided by studies of vole cycles in Europe (Hanski *et al.*, 1991; see also Section 14.6.4). In subarctic Finnish Lapland, there are regular 4- or 5-year cycles, with a ratio of maximum : minimum vole densities generally exceeding 100. In southern Sweden small rodents show no regular multiannual cycles. But between the two, moving north to south in Fennoscandia, there is a gradient of decreasing regularity, amplitude and length of the cycle. Hanski *et al.* argue that this gradient is itself correlated with a gradient of increasing densities of generalist predators that switch between alternative

switching, stabilization and the voles of Fennoscandia

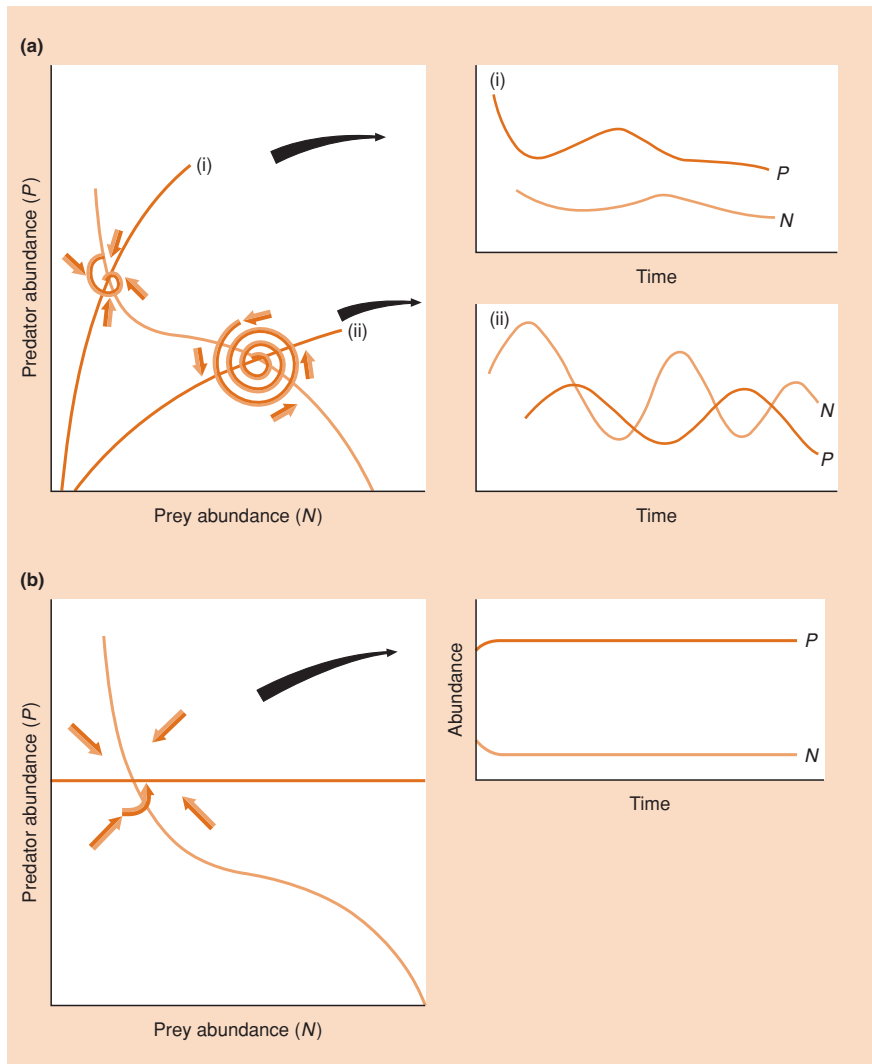


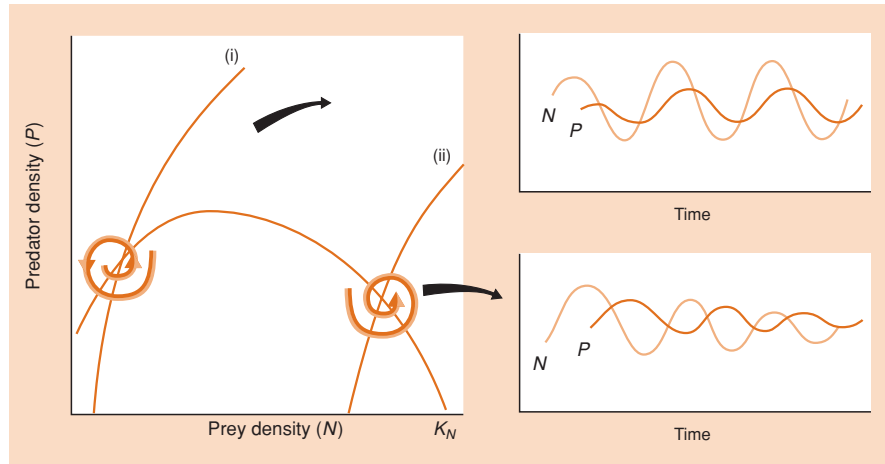
Figure 10.11 (a) The prey zero isocline is that which is appropriate when consumption rate is particularly low at low prey densities because of a type 3 functional response, an aggregative response (and partial refuge), an actual refuge or because of a reserve of plant material that is not palatable. With a relatively inefficient predator, predator zero isocline (ii) is appropriate and the outcome is not dissimilar from Figure 10.7. However, a relatively efficient predator will still be able to maintain itself at low prey densities. Predator zero isocline (i) will therefore be appropriate, leading to a stable pattern of abundance in which prey density is well below the carrying capacity and predator density is relatively high. (b) When a type 3 functional response arises because the predator exhibits switching behavior, the predator's abundance may be independent of the density of any particular prey type (main figure), and the predator zero isocline may therefore be horizontal (unchanging with prey density). This can lead to a stable pattern of abundance (inset) with prey density well below the carrying capacity.

prey as relative densities change (especially red foxes, badgers, domestic cats, buzzards, tawny owls and crows) and of specialist bird predators (especially other owl species and kestrels) that, being wide ranging in their activity, switch between alternative areas. In both cases, predator dynamics would be effectively independent of vole abundance, adding stability to the system in the manner of Figure 10.11b. In fact, Hanski *et al.* were able to go further in constructing a simple model of prey (voles) interacting with specialist predators (mustelids: stoats and weasels) and generalist (switching) predators. Their general contention was supported; as the number of generalist predators increased, oscillations in vole and mustelid abundance (which may or may not be the basis for the vole cycle) decreased in length and amplitude. Large enough densities of switching generalists stabilized the cycle entirely.

Turning to type 2 responses, if the predator has a response that reaches its plateau at relatively low prey densities (well below K_N), then the prey zero isocline has a hump, because there is a range of intermediate prey densities where the predators become less efficient with increasing prey density but the effects of competition amongst the prey are not intense. A hump will also arise here if the prey are subject to an 'Allee effect', where they have a disproportionately low rate of recruitment when their own density is low, perhaps because mates are difficult to find or because a 'critical number' must be exceeded before a resource can be properly exploited, i.e. there is inverse density dependence at low population densities (Courchamp *et al.*, 1999). If the predator

type 2 responses and the Allee effect destabilize – but not necessarily in practice

Figure 10.12 The possible effects of a prey isocline with a 'hump', either as a result of a type 2 functional response or an Allee effect. (i) If the predator is highly efficient, with its isocline crossing to the left of the hump, then the hump can be destabilizing, leading to the persistent oscillations of a limit cycle (inset). (ii) But if the predator is less efficient, crossing to the right of the hump, then the hump has little effect on the dynamics: the oscillations converge (inset).



isocline crosses to the right of the hump, then the population dynamics of the interaction will be little affected; but if the isocline crosses to the left of the hump, then the outcome will be persistent rather than convergent oscillations, i.e. the interaction will be destabilized (Figure 10.12).

However, for a type 2 response to have this effect, predators would have to suffer serious reductions in their consumption rate at prey densities far below those at which the prey themselves suffer seriously from competition. This is unlikely. The potentially destabilizing effects of type 2 responses may also therefore be of little practical importance.

A destabilizing Allee effect has not apparently been established for any 'natural' predator-prey interaction. On the other hand, when we ourselves are the predator (for example, with exploited fisheries populations), we frequently have the ability (i.e. the technology) to maintain effective predation at low prey densities. If the prey population also exhibits an Allee effect, then the combination of this and persistent predation may all too readily drive a population towards extinction (Stephens & Sutherland, 1999; and see Section 15.3.5). That is, our isocline may cross that of the prey well to the left of their hump.

10.5 Heterogeneity, aggregation and spatial variation

Until now in this chapter, environmental heterogeneities, and the variable responses of predators and prey to such heterogeneities – all of which we saw in the previous chapter to be commonplace – have been ignored. We can ignore them no longer.

10.5.1 Aggregative responses to prey density

Because of the potential consequences for population dynamics, ecologists have been particularly interested in patch preferences

where patches vary in the density of the food or prey items they contain (see Section 9.6). At one time it appeared, and was widely believed, that: (i) predators generally spent most time in patches containing high densities of prey (because these were the most profitable patches); (ii) most predators were therefore to be found in such patches; and (iii) prey in those patches were therefore most vulnerable to predation, whereas those in low-density patches were relatively protected and most likely to survive. Examples certainly exist to support the first two of these propositions (see Figure 9.20a–d), demonstrating an 'aggregative response' by the predators that is directly density dependent (predators spending most time in patches with high densities of prey such that prey and predator densities are positively correlated). However, this is not always the case. Furthermore, contrary to the third proposition, reviews of host-parasitoid interactions (e.g. Pacala & Hassell, 1991) have shown that prey (hosts) in high-density patches are not necessarily the most vulnerable to attack (direct density dependence): percentage parasitism may also be inversely density dependent or density independent between patches (see Figure 9.20e). Indeed, the reviews suggest that only around 50% of the studies examined show evidence of density dependence, and in only around 50% of these is the density dependence direct, as opposed to inverse. None the less, despite this variation in pattern, it remains true that the risk of predation often varies greatly between patches, and hence between individual prey.

Many herbivores also display a marked tendency to aggregate, and many plants show marked variation in their risk of being attacked. The cabbage aphid (*Brevicoryne brassicae*) forms aggregates at two separate levels (Way & Cammell, 1970). Nymphs quickly form large groups when isolated on the surface of a single leaf, and populations on a single plant tend to be restricted to particular leaves. When aphids attack only one leaf

do predators aggregate in high-density prey patches?

plants may be protected by the aggregative responses of herbivores

of a four-leaved cabbage plant (as they do naturally), the other three leaves survive; but if the same number of aphids are evenly spread over the four leaves, then all four leaves are destroyed (Way & Cammell, 1970). The aggregative behavior of the herbivores affords protection to the plant overall. But how might such heterogeneities influence the dynamics of predator-prey interactions?

10.5.2 Heterogeneity in the graphical model

refuges, partial
refuges and vertical
isoclines

We can start by incorporating into the Lotka–Volterra isoclines some relatively simple types of heterogeneity.

Suppose that a portion of the prey population exists in a refuge: for example, shore snails packed into cracks in the cliff-face, away from marauding birds, or plants that maintain a reserve of material underground that cannot be grazed. In such cases, the prey zero isocline rises vertically at low prey densities (again, see Figure 10.11), since prey at low densities, hidden in their refuge, can increase in abundance irrespective of predator density.

Even if predators tend simply to ignore prey in low-density patches, as we have seen in some aggregative responses (see Section 9.6), this comes close to those prey being in a refuge, in the sense that the predators do not (rather than cannot) attack them. The prey may therefore be said to have a ‘partial refuge’, and this time the prey isocline can be expected to rise almost vertically at low prey abundances.

We saw above, when discussing type 3 functional responses, that such isoclines have a tendency to stabilize interactions. Early analyses of both the Lotka–Volterra and the Nicholson–Bailey systems (and early editions of this textbook) agreed with this conclusion: that spatial heterogeneities, and the responses of predators and prey to them, stabilize predator–prey dynamics, often at low prey densities (Beddington *et al.*, 1978). However, as we shall see next, subsequent developments have shown that the effects of heterogeneity are more complex than was previously supposed: the effects of heterogeneity vary depending on the type of predator, the type of heterogeneity, and so on.

10.5.3 Heterogeneity in the Nicholson–Bailey model

negative binomial
encounters . . .

Most progress has been made in untangling these effects in host–parasitoid systems. A good starting

point is the model constructed by May (1978), in which he ignored precise details and argued simply that the distribution of host–parasitoid encounters was not random but aggregated. In particular, he assumed that this distribution could be described by a particular statistical model, the negative binomial. In this case

(in contrast to Section 10.2.3), the proportion of hosts not encountered at all is given by:

$$p_0 = \left[1 + \frac{AP_t}{k} \right]^{-k} \quad (10.20)$$

where k is a measure of the degree of aggregation; maximal aggregation at $k = 0$, but a random distribution (recovery of the Nicholson–Bailey model) at $k = \infty$. If this is incorporated into the Nicholson–Bailey model (Equations 10.14 and 10.15), then we have:

$$H_{t+1} = H_t e^r \left[1 + \frac{AP_t}{k} \right]^{-k} \quad (10.21)$$

$$P_{t+1} = H_t \left\{ 1 - \left[1 + \frac{AP_t}{k} \right]^{-k} \right\}. \quad (10.22)$$

The behavior of a version of this model, which also includes a density-dependent host rate of increase, is illustrated in Figure 10.13, from which it is clear that the system is given a marked boost in stability by the incorporation of significant levels of aggregation ($k \leq 1$). Of particular importance is the existence of stable systems with low values of H^*/K ; i.e. aggregation appears capable of generating stable host abundances well below the host’s normal carrying capacity. This coincides with the conclusion drawn from Figure 10.11.

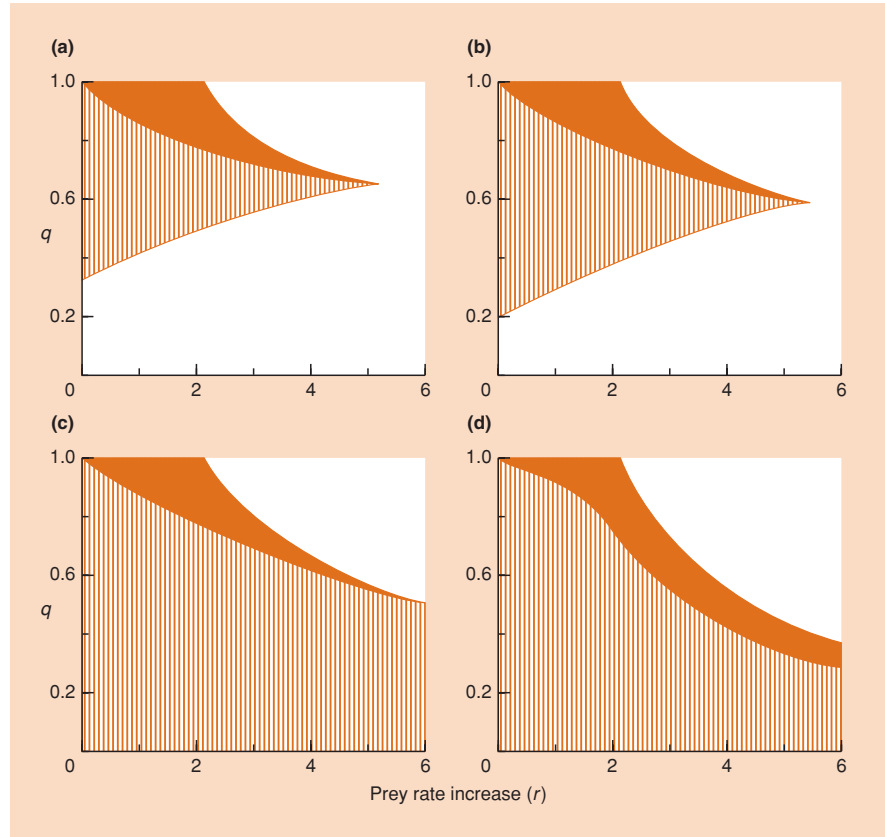
. . . that stabilize
dynamics

10.5.4 Aggregation of risk and spatial density dependence

How does this stability arise out of aggregation? The answer lies in what has been called ‘pseudo-interference’ (Free *et al.*, 1977). With mutual interference, as predator density increases, predators spend an increasing amount of time interacting with one another, and their attack rate therefore declines. With pseudo-interference, attack rate also declines with parasitoid density, but as a result of an increasing fraction of encounters being wasted on hosts that have already been attacked. The crucial point is that ‘aggregation of risk’ amongst hosts tends to increase the amount of pseudo-interference. At low parasitoid densities, a parasitoid is unlikely to have its attack rate reduced as a result of aggregation. But at higher parasitoid densities, parasitoids in aggregations (where most of them are) will increasingly be faced with host patches in which most or all of the hosts have already been parasitized. As parasitoid density increases, therefore, their effective attack rate (and hence their subsequent birth rate) declines rapidly – a *directly* density-dependent effect. This dampens both the natural oscillations in parasitoid density, and their impact on host mortality.

pseudo-interference

Figure 10.13 May's (1978) model of host–parasitoid aggregation, with host self-limitation incorporated, illustrates that aggregation can enhance stability and give rise to stability at low values of $q = H^*/K$. In the solid orange area there is an exponential approach to equilibrium; in the hatched area there is an oscillatory approach to equilibrium; outside these there is instability (the oscillations either diverge or are sustained). The four figures are for four values of k , the exponent of the negative binomial distribution in the model: (a) $k = \infty$: no aggregation, least stability; (b) $k = 2$; (c) $k = 1$; (d) $k = 0.1$: most aggregation. (After Hassell, 1978.)



aggregation of risk strengthens direct (temporal) density dependences

To summarize, aggregation of risk stabilizes host–parasitoid interactions by strengthening direct (not delayed) density dependencies that already exist (Taylor, 1993). The stabilizing powers of this spatial phenomenon, aggregation of risk, therefore arise not from any spatial density dependencies, but from its translation into direct, temporal density dependence.

But how does aggregation of risk relate to the aggregative responses of parasitoids? And do aggregative responses and aggregation of risk necessarily lead to enhanced stability? We can address these questions by examining Figure 10.14, bearing in mind from Section 9.6 that aggregated predators do not necessarily spend most time foraging in patches of high host density (spatial density dependence); foraging time can also be negatively correlated with host density (inverse density dependence) or independent of host density. Start with Figure 10.14a. The distribution of parasitoids over host patches follows a perfect, straight line density-dependent relationship. But, since the host : parasitoid ratio is therefore the same in each host patch, the risk is likely to be the same in each host patch, too. Thus, positive spatial density dependence does not necessarily lead to aggregation of risk and does not necessarily enhance stability. On the other hand, with a directly density-dependent relationship that

accelerates (Figure 10.14b), there does appear to be aggregation of risk, and this might well enhance stability (Hassell & May, 1973); but it turns out that whether or not it does so depends on the parasitoids' functional response (Ives, 1992a). With a type 1 response, assumed by most analyses, stability is enhanced. But with a more realistic type 2 response, initial increases in density-dependent aggregation from zero aggregation actually decrease aggregation of risk and are destabilizing. Only high levels of density-dependent aggregation are stabilizing.

Moreover, it is clear from Figures 10.14c and d that there can be considerable aggregation of risk with either inverse spatial density dependence or no spatial density dependence of any sort – and these would not be counteracted by a type 2 functional response. In partial answer to our two questions above, therefore, aggregative responses that are spatially density dependent are actually *least* likely to lead to aggregation of risk, and therefore least likely to enhance stability.

In practice, of course, with real sets of data (like those in Figure 9.20), aggregation of risk will often arise from a combination of spatially density-dependent (direct or inverse) and density-independent responses (Chesson & Murdoch, 1986; Pacala & Hassell, 1991). Pacala, Hassell and coworkers have

aggregative responses and aggregation of risk

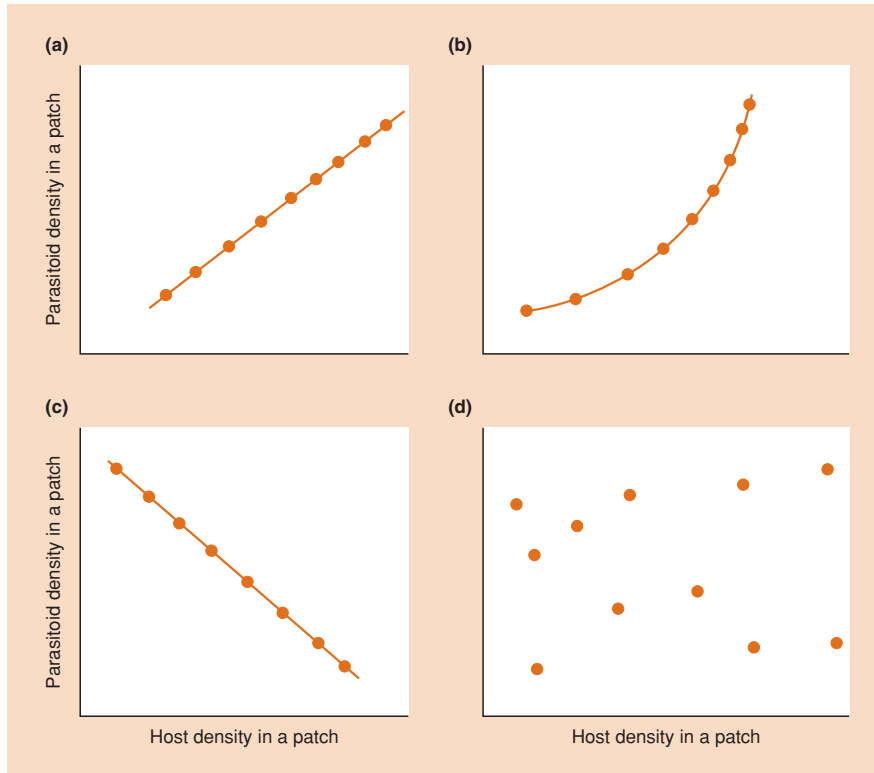


Figure 10.14 The aggregative responses of parasitoids and the aggregation of risk. (a) Parasitoids aggregate in high host-density patches, but the parasitoid : host ratio is the same in all patches (a perfect straight-line relationship), and hence the risk to hosts is apparently the same in all patches. (b) Parasitoid aggregation to high host-density patches now accelerates with increasing host density, and hosts in high-density patches are thus apparently at greater risk of parasitization: there is aggregation of risk. (c) With perfect inverse density dependence (i.e. parasitoid aggregation in low host-density patches) the hosts in the low-density patches are apparently at a much greater risk of parasitization: again there is aggregation of risk. (d) Even with no aggregative response (density independence) the hosts in some patches are apparently at a greater risk of parasitization (are subject to a higher parasitoid : host ratio) than others: here too there is aggregation of risk.

called the former the ‘host density dependent’ (HDD) component, and the latter the ‘host density independent’ (HDI) component, and have described methods by which, in real data sets like Figure 9.20, the aggregation of risk can be split between them. Interestingly, in an analysis of 65 data sets, representing 26 different host–parasitoid combinations (Pacala & Hassell, 1991), 18 appeared to have sufficient aggregation of risk to stabilize their interactions, but for 14 of these 18 cases, it was HDI variation that contributed most to the total, further weakening any imagined link between spatial density dependence and stability.

10.5.5 Heterogeneity in some continuous-time models

We have been pursuing parasitoids and hosts; and in doing so we have been retaining certain structural features in our analysis that we should now reconsider. In particular, our parasitoids have been assumed, in effect, to arrange themselves over host patches at the beginning of a generation (or whatever the time interval is between t and $t + 1$), and then to have to suffer the consequences of that arrangement until the beginning of the next generation. But suppose we move into continuous time – as appropriate for many parasitoids as it is for many other predators. Now, aggregation should be assumed to occur on a continuous basis, too. Predators in a depleted, or even a depleting, patch should leave

and redistribute themselves (see Section 9.6.2). The whole basis of pseudo-interference and hence stability, namely wasted predator attacks in high predator density patches, tends to disappear.

Murdoch and Stewart-Oaten (1989)

went to, perhaps, the opposite extreme to the one we have been considering, by constructing a continuous-time model in which prey moved instantly into

continuous
redistribution of
predators and prey

patches to replace prey that had been consumed, and predators moved instantly into patches to maintain a consistent pattern of predator–prey covariation over space. The effects on their otherwise neutrally stable Lotka–Volterra model contrast strongly with those we have seen previously. First, predator aggregation that is independent of local prey density now has *no* effect on either stability or prey density. However, predator aggregation that is directly dependent on local prey density has an effect that depends on the strength of this dependence – although it always lowers prey density (because predator efficiency is increased). If such density dependence is relatively weak (as Murdoch and Stewart-Oaten argue it usually is in practice), then stability is *decreased*. Only if it is stronger than seems typical in nature is stability increased.

Other, less ‘extreme’ continuous-time formulations (Ives, 1992b), or those that combine discrete generations with redistribution within generations (Rohani *et al.*, 1994), produce results

that are themselves intermediate between the ‘Nicholson–Bailey extreme’ and the ‘Murdoch–Stewart-Oaten extreme’. It seems certain, however, that a preoccupation with models lacking within-generation movement has, in the past, led to a serious overestimation of the significance of aggregation to patches of high host density in stabilizing host–parasitoid interactions.

10.5.6 The metapopulation perspective

The continuous- and discrete-time approaches clearly differ, but they share a common perspective in seeing predator–prey interactions occurring within single populations, albeit populations with inbuilt variability. An alternative is a ‘metapopulation’ perspective (see Section 6.9), in which environmental patches support subpopulations that have their own internal dynamics, but are linked to other subpopulations by movement between patches.

A number of studies have investigated predator–prey metapopulation models, usually with unstable dynamics within patches. Mathematical difficulties have often limited analysis to two-patch models, where, if the patches are the same, and dispersal is uniform, stability is unaffected: patchiness and dispersal have no effect in their own right (Murdoch *et al.*, 1992; Holt & Hassell, 1993).

patch differences stabilize through asynchrony

Differences between the patches, however, tend, in themselves, to stabilize the interaction (Ives, 1992b; Murdoch *et al.*, 1992; Holt & Hassell, 1993). The reason is that any difference in parameter values between patches leads to asynchrony in the fluctuations in the patches. Inevitably, therefore, a population at the peak of its cycle tends to lose more by dispersal than it gains, a population at a trough tends to gain more than it loses, and so

on. Dispersal and asynchrony together, therefore, give rise to stabilizing temporal density dependence in net migration rates.

The situation becomes much more complex with the inclusion of aggregative behavior, since dispersal rates themselves become a much more complex function of both prey and predator densities. Aggregation appears to have two opposing effects (Murdoch *et al.*, 1992). It tends to increase the asynchrony between fluctuations in predator abundance (enhancing stability) but to reduce the asynchrony between prey fluctuations (decreasing stability). The balance between these forces appears to be sensitive to the strength of the aggregation, but perhaps even more sensitive to the assumptions built into the models (Godfray & Pacala, 1992; Ives, 1992b; Murdoch *et al.*, 1992). Aggregation may either stabilize or destabilize. In contrast to previous analyses, it has no clear effect on prey density since its stabilizing powers are not linked to predator efficiency.

The treatment of a spatially heterogeneous predator–prey interaction as a problem in metapopulation dynamics was taken a stage further by Comins *et al.* (1992). They constructed computer models of an environment consisting of a patchwork of squares, which could actually be visualized as such (Figure 10.15). In each generation, two processes occurred in sequence. First, a fraction, μ_p , of predators, and a fraction, μ_N , of prey, dispersed from each square to the eight neighboring squares. At the same time, predators and prey from the eight neighboring squares were dispersing into the first square. Thus, for example, the dynamics for the density of prey, $N_{i,t+1}$, in square i in generation $t + 1$, was given by:

$$N_{i,t+1} = N_{i,t}(1 - \mu_N) + \mu_N \bar{N}_{i,t} \tag{10.23}$$

an explicitly, and visually, spatial model

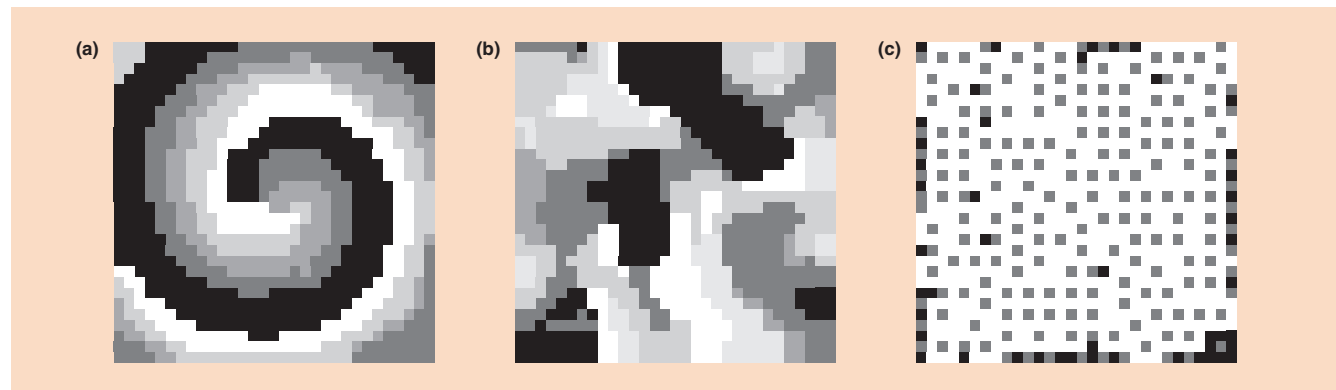


Figure 10.15 Instantaneous maps of population density for simulations of the dispersal model of Comins *et al.* (1992) with Nicholson–Bailey local dynamics. Different levels of shading represent different densities of hosts and parasitoids. Black squares represent empty patches; dark shades becoming paler represent patches with increasing host densities; light shades to white represent patches with hosts and increasing parasitoid densities. (a) Spirals: $\mu_N = 1$, $\mu_p = 0.89$; (b) spatial chaos: $\mu_i = 0.2$, $\mu_p = 0.89$; (c) a ‘crystalline lattice’: $\mu_N = 0.05$; $\mu_p = 1$. (After Comins *et al.*, 1992.)

or:

$$N_{i,t+1} = N_{i,t} + \mu_N (\bar{N}_{i,t} - N_{i,t}), \quad (10.24)$$

where $\bar{N}_{i,t}$ is the mean density in the eight squares neighboring square i in generation t . The second phase then consisted of one generation of standard predator–prey dynamics, either following the Nicholson–Bailey equations or a discrete-time version of the Lotka–Volterra equations (May, 1973). Simulations were started with random prey and predator populations in a single patch, with all the other patches empty.

We know that within individual squares, if they existed in isolation, the dynamics would be unstable. But within the patchwork of squares as a whole, stable or at least highly persistent patterns can readily be generated (Figure 10.15). The general message is similar to the results that we have already seen: that stability can be generated by dispersal in metapopulations in which different patches are fluctuating asynchronously. Note especially, in this case, that a patch experiences a net gain in migrants when its density is lower than the mean of the eight patches with which it connects (Equation 10.24) but experiences a net loss when its density is higher – a kind of density dependence. Note, too, that the asynchrony arises in the present case because the population has spread from a single initial patch (all patches are, in principle, the same) and that it is maintained by dispersal being limited to the neighboring patches (rather than being a powerful equalizing force over all patches).

emergent spatial patterns

Moreover, the explicitly spatial aspects of this model have, quite literally, added another dimension to the results. Depending on the dispersal fractions and the host reproductive rate, a number of quite different spatial structures can be generated (although they tend to blur into one another) (Figure 10.15a–c). ‘Spatial chaos’ can occur, in which a complex set of interacting wave fronts are established, each one persisting only briefly. With somewhat different parameter values, and especially when both predator and prey are highly mobile, the patterns are more structured than chaotic, with ‘spiral waves’ rotating around almost immobile focal points. The model, therefore, makes the point very graphically that persistence at the level of a whole population does not necessarily imply either uniformity across the population or stability in individual parts of it. Static ‘crystalline lattices’ can even occur within a narrow range of parameter values, with highly mobile predators and rather sedentary prey, emphasizing that pattern can be generated internally within a population even in an intrinsically homogeneous environment.

Is there one general message that can be taken from this body of theory? Certainly, we cannot say ‘aggregation does *this* or *that* to predator–prey interactions’. Rather, aggregation can have a variety of effects, and knowing which of these is likely will require detailed knowledge of predator and prey biology for the inter-

action concerned. In particular, the effects of aggregation have been seen to depend on the predator’s functional response, the extent of host self-regulation, and so on – other features that we have examined in isolation. It is necessary, as stressed at the beginning of this chapter, in seeking to understand complex processes, to isolate conceptually the different components. But it is also necessary, ultimately, to recombine those components.

10.5.7 Aggregation, heterogeneity and spatial variation in practice

What, then, can be said about the role of spatial variation in practice? The stabilizing effects of heterogeneity were demonstrated famously, long ago, by Huffaker (1958; Huffaker *et al.*, 1963), who studied a system in which a predatory mite fed on a herbivorous mite, which fed on oranges interspersed amongst rubber balls in a tray. In the absence of its predator, the prey maintained a fluctuating but persistent population (Figure 10.16a); but if the predator was added during the early stages of prey population growth, it rapidly increased its own population size, consumed all of its prey and then became extinct itself (Figure 10.16b). The interaction was altered, however, when Huffaker made his microcosm more ‘patchy’ (creating, effectively, a metapopulation, though the term had not been coined at the time). He spread the oranges further apart, and partially isolated each one by placing a complex arrangement of vaseline barriers in the tray, which the mites could not cross. But he facilitated the dispersal of the prey by inserting a number of upright sticks from which they could launch themselves on silken strands carried by air currents. Dispersal between patches was therefore much easier for the prey than it was for the predators. In a patch occupied by both predators and prey, the predators consumed all the prey and then either became extinct themselves or dispersed (with a low rate of success) to a new patch. But in patches occupied by prey alone, there was rapid, unhampered growth accompanied by successful dispersal to new patches. In a patch occupied by predators alone, there was usually death of the predators before their food arrived. Each patch was therefore ultimately doomed to the extinction of both predators and prey. But overall, at any one time, there was a mosaic of unoccupied patches, prey–predator patches heading for extinction, and thriving prey patches; and this mosaic was capable of maintaining persistent populations of both predators and prey (Figure 10.16c).

Subsequently, others, too, have demonstrated the power of a metapopulation structure in promoting the persistence of coupled predator and prey populations when their dynamics

metapopulation effects in mites, beetles and ciliates

in individual subpopulations are unstable. Figure 10.17a, for example, shows this for a parasitoid attacking its beetle host. Figure 10.17b shows similar results for prey and predatory ciliates

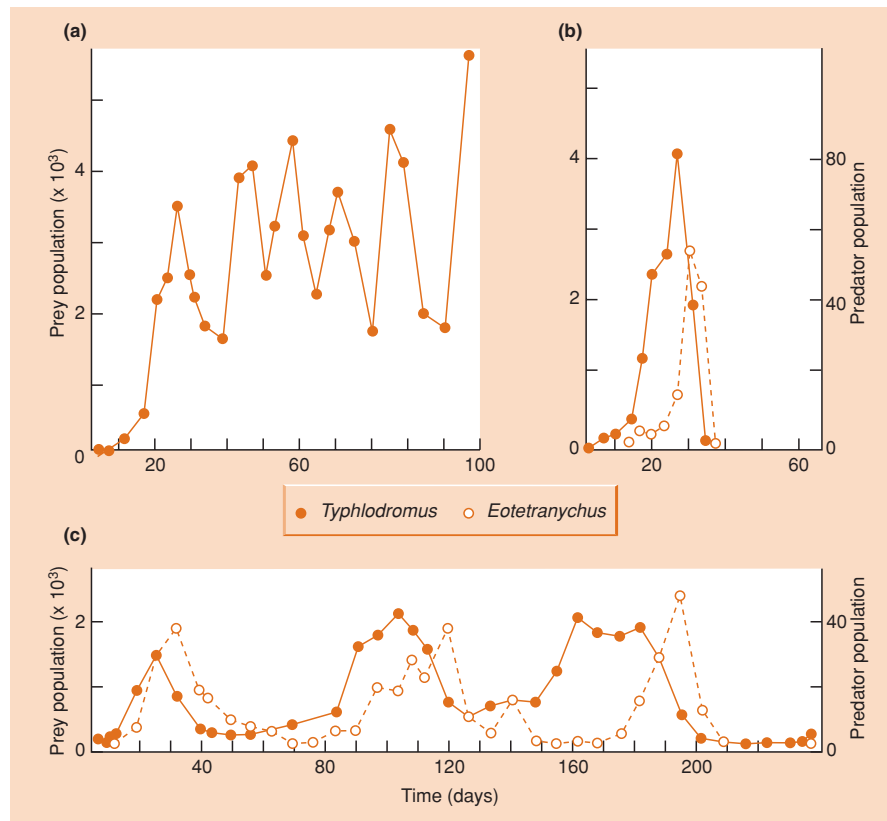


Figure 10.16 Hide and seek: predator-prey interactions between the mite *Eotetranychus sexmaculatus* and its predator, the mite *Typhlodromus occidentalis*. (a) Population fluctuations of *Eotetranychus* without its predator. (b) A single oscillation of the predator and prey in a simple system. (c) Sustained oscillations in a more complex system. (After Huffaker, 1958.)

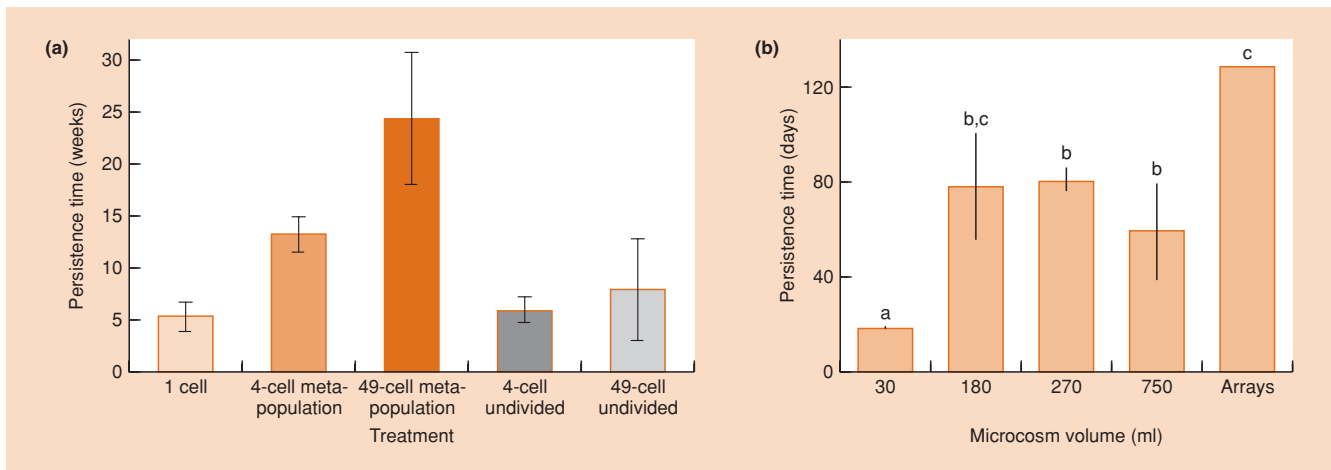


Figure 10.17 A metapopulation structure can increase the persistence of predator-prey interactions. (a) The parasitoid, *Anisopteromalus calandrae*, attacking its bruchid beetle host, *Callosobruchus chinensis*, living on beans either in small single 'cells' (short persistence time, left), or in combinations of cells (four or 49), which either had free access between them so that they effectively constituted a single population (persistence time not significantly increased, right), or had limited (infrequent) movement between cells so that they constituted a metapopulation of separate subpopulations (increased persistence time, center). Bars show standard errors. (After Bonsall *et al.*, 2002.) (b) The predatory ciliate, *Didinium nasutum*, feeding on the bacterivorous ciliate, *Colpidium striatum*, in bottles of various volumes, where persistence time varied little, except in the smallest populations (30 ml) where times were shorter, and also in 'arrays' of nine or 25 linked 30 ml bottles (metapopulations), where persistence was greatly prolonged: all populations persisted until the end of the experiment (130 days). Bars show standard errors; different letters above bars indicate treatments that were significantly different from one another ($P < 0.05$). (After Holyoak & Lawler, 1996.)

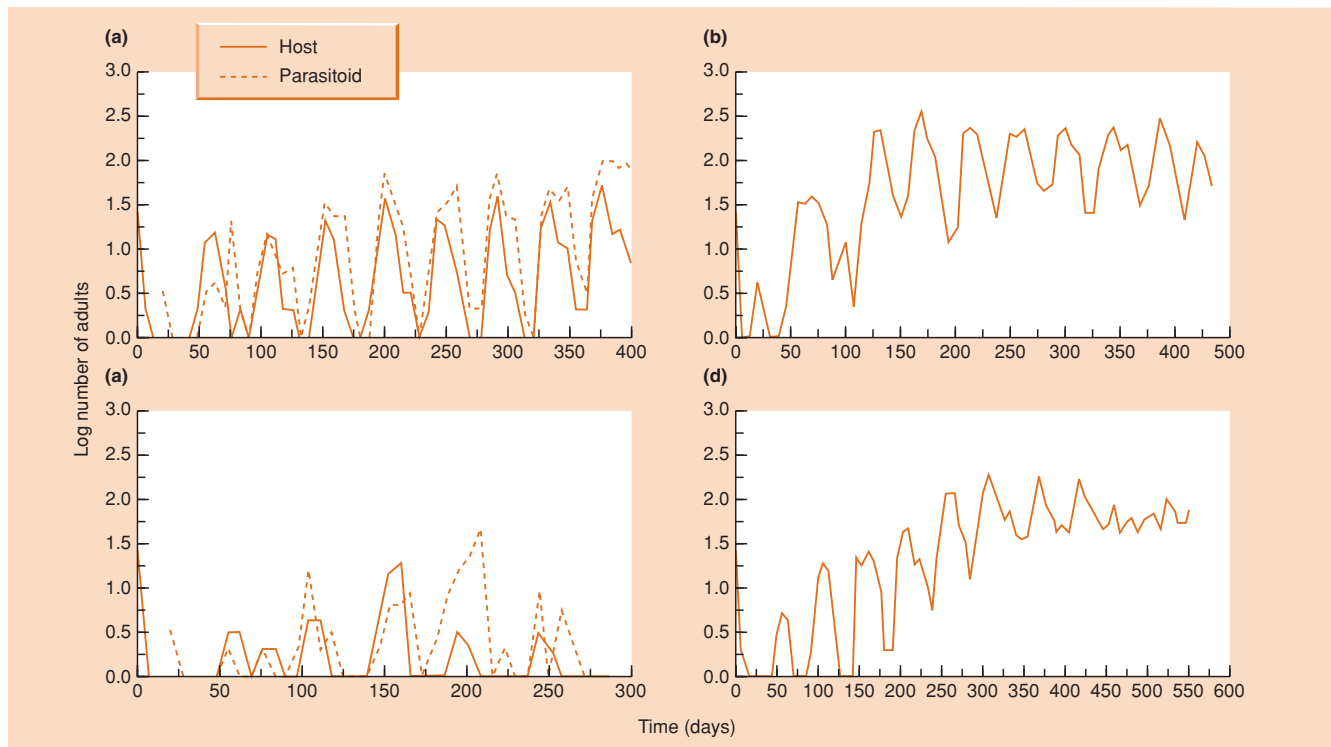


Figure 10.18 Long-term population dynamics in laboratory population cages of a host (*Plodia interpunctella*), with and without its parasitoid (*Venturia canescens*). (a) Host and parasitoid in deep medium, exhibiting coupled cycles in abundance, approximately one host generation in length. (b) The host alone in deep medium, exhibiting similar cycles. (c) Host and parasitoid in shallow medium, unable to persist. (d) The host alone in shallow medium, able to persist. The deep medium provides a refuge from attack for a proportion of the host population that is not present in the shallow medium (see Section 10.5.2). All data sets are selected from several replicates showing the same pattern. (After Begon *et al.*, 1995.)

(protists), where, in support of the role of a metapopulation structure, it was also possible to demonstrate asynchrony in the dynamics of individual subpopulations and frequent local prey extinctions and recolonizations (Holyoak & Lawler, 1996).

a refuge for a moth

A study providing support for the stabilizing powers of a physical refuge is illustrated in Figure 10.18, based on the same *Plodia*–*Venturia* host–parasitoid system as that in Figure 10.1c. In this case, hosts living deeper in their food are beyond the reach of the parasitoids attempting to lay their eggs in them. In the absence of this refuge, in a shallow food medium, this host–parasitoid interaction is unable to persist (Figure 10.18c), although the host alone does so readily (Figure 10.18d). With a refuge present, however, in a deeper food medium, the host and parasitoid can apparently persist together indefinitely (Figure 10.18a).

more mites: a metapopulation or a refuge?

In fact, though, the distinctions between different types of spatial heterogeneity may not be as clear cut in real systems as they are in mathematical models. Ellner *et al.* (2001), for example, examined a system of predatory mites, *Phytoseiulus*

persimilis, feeding on herbivorous mites, *Tetranychus urticae*, feeding on bean plants, *Phaseolus lunatus*. On individual plants and on a single ‘continent’ of 90 plants (Figure 10.19a), the system had no long-term persistence (Figure 10.19c). However, when the Styrofoam sheet supporting the plants was split into eight islands of 10 plants, connected by bridges that limited the mites’ powers of dispersal (Figure 10.19b), persistence was apparently indefinite (Figure 10.19d, e). It would be easy to jump to the conclusion that stability was increased by the eight-island metapopulation structure. But when Ellner *et al.* examined mathematical models of the system that allowed the various aspects of the altered layout to be investigated one by one, they could detect no significant effect of such a structure. Instead, they suggested that the enhanced stability arose from a different aspect: a reduction in the predators’ ability to detect and respond to prey outbreaks on individual plants – a prey ‘refuge’ effect that could arise in the absence of any explicit spatial structure.

One major problem in making pronouncements about the stabilizing role of aggregation of risk is that although, as we have seen, there have been wide-

real data confirm the complexity of natural systems

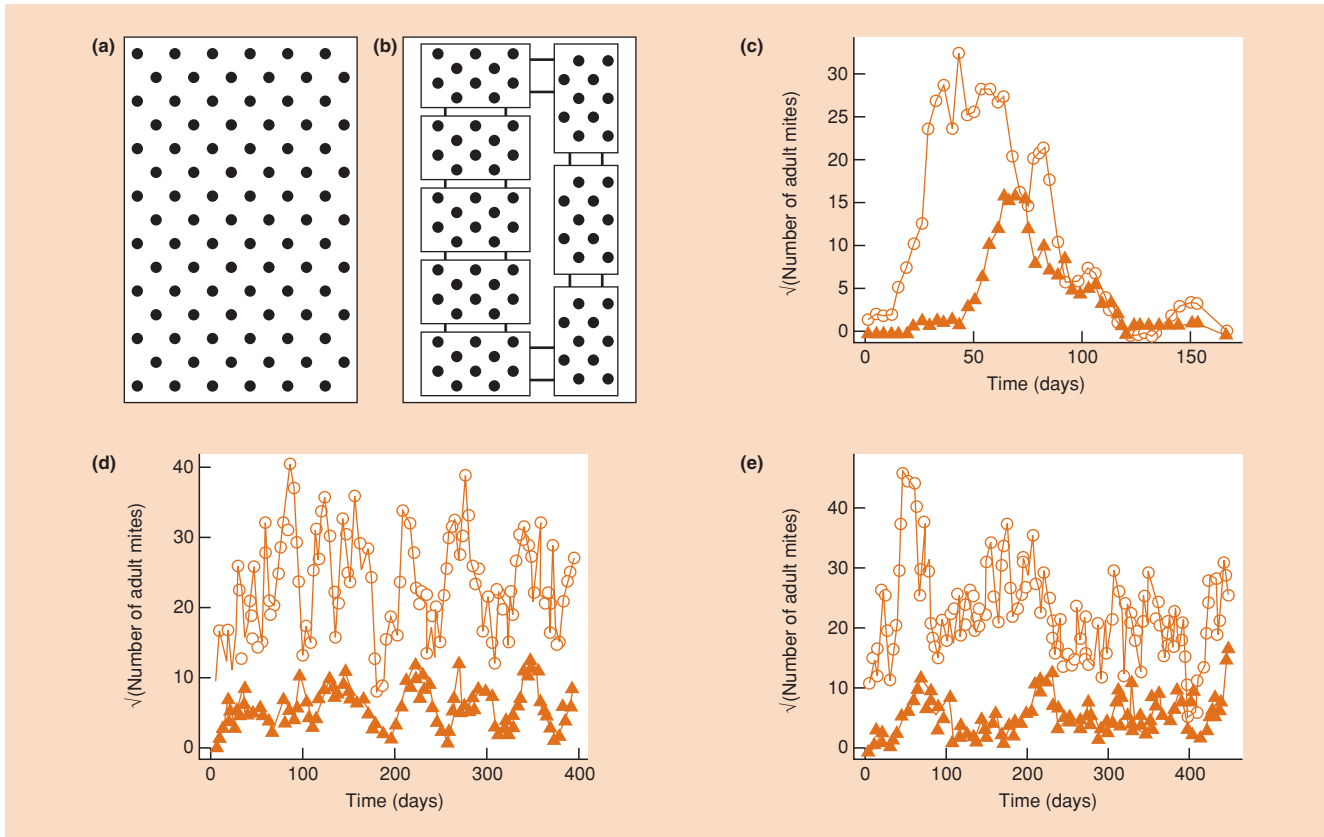


Figure 10.19 The population dynamics of the predatory mite, *Phytoseiulus persimilis*, and its herbivorous mite prey, *Tetranychus urticae*. They interacted either (a) on a single continent of 90 bean plants, the dynamics of which are shown in (c) (\blacktriangle , predators; \circ , prey), or (b) in a metapopulation of eight islands of 10 plants. For the latter, the dynamics of two replicates are shown in (d) and (e), where persistence (stability) is clearly enhanced. (After Ellner *et al.*, 2001.)

ranging surveys of the data on spatial distributions of attacks, these data generally come from studies of very short duration – often of only one generation. We do not know if the observed spatial patterns are typical for that interaction; nor do we know if the population dynamics show the degree of stability that the spatial patterns might seem to predict. One investigation that has examined population dynamics and spatial distributions over several generations is that of Redfern *et al.* (1992), who made a 7-year (seven-generation) study of two tephritid fly species that attack thistles and the guilds of parasitoids that attack the flies. For one host, *Terellia serratulae* (Figure 10.20a) there was evidence of year-to-year density dependence in the overall rate of parasitism (Figure 10.20b), but no strong evidence of significant levels of aggregation within generations, either overall (Figure 10.20c) or for parasitoid species individually. For the other species, *Urophora stylata* (Figure 10.20d), there was no apparent temporal density dependence but good evidence for the aggregation of risk (Figure 10.20e, f), and, to repeat a pattern we have seen before, most heterogeneity was contributed by the HDI component. It cannot be said, however, that the patterns of this study fit neatly,

overall, to the theory we have outlined. First, both hosts were attacked by several parasitoid species – not one, as assumed by most models. Second, the levels of aggregation (and to a lesser extent the HDI or HDD contributions) varied considerably and apparently randomly from year to year (Figure 10.20c, f): no one year was typical, and no single ‘snap-shot’ could have captured either interaction. Finally, while the relatively stable dynamics of *Terellia* may have reflected the more demonstrable direct density dependence in parasitism, this appeared to be quite unconnected to any differences in the aggregation of risk.

The effects of spatial heterogeneities on the stability of predator–prey dynamics are not only of purely scientific interest. They have also been the subject of lively debate (Hawkins & Cornell, 1999) in considering the properties and nature of biological control agents: natural enemies of a pest that are imported into an area, or otherwise aided and abetted, in order to control the pest (see Section 15.2.5). What is required of a good biological control agent is the ability to

spatial heterogeneity and the most effective biological control agents

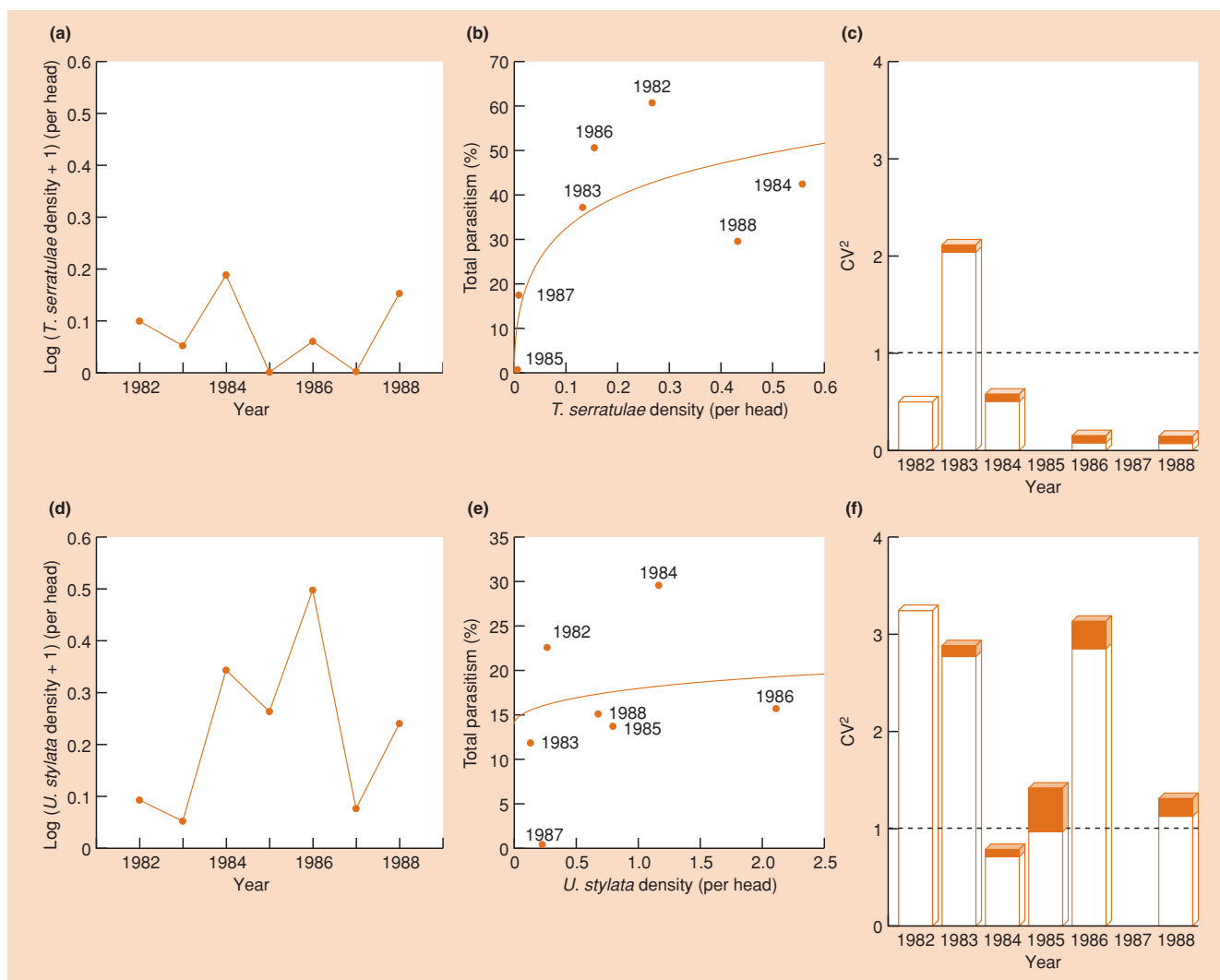


Figure 10.20 Attacks by parasitoids on tephrytid flies (*Terellia serratulae* and *Urophora stylata*) that attack thistle flower-heads. The dynamics of the populations are shown for *T. serratulae* in (a) and for *U. stylata* in (d). Temporal density dependence of parasitoid attacks on *T. serratulae* (b) is significant ($r^2 = 0.75$; $P < 0.05$), but for *U. stylata* (e) it is not ($r^2 = 0.44$; $P > 0.05$); both fitted lines take the form $y = a + b \log_{10} x$. However, whereas for *T. serratulae* (c) there is little aggregation of risk of parasitoid attack within years (measured as $CV^2 > 1$ for aggregation), with *U. stylata* (f) there is far more, most of which is HDI (no shading) rather than HDD (dark shading). (After Redfern *et al.*, 1992.)

reduce the prey (pest) to a stable abundance well below its normal, harmful level, and we have seen that some theoretical analyses suggest that this is precisely what aggregative responses help to generate. Establishing such a link in practice, however, has not proved easy. Murdoch *et al.* (1995), for example, noted that the California red scale, *Aonidiella aurantii*, an insect pest of citrus plants worldwide, appeared to be kept at low and remarkably stable densities in southern California by a parasitoid introduced to control it, *Aphytis mellitus*. The existence of a partial refuge from parasitization for the red scale seemed a plausible hypothesis for how this was achieved: on bark in the interior of the

trees, rates of parasitism were very low and scale densities high, seemingly as a result of the activities of ants there that interfered with the searching parasitoids. Murdoch *et al.*, therefore, tested this hypothesis by a field experiment in which ants were removed from a number of trees. Parasitization rates in the refuge did increase, and scale abundance declined there (Figure 10.21), and there was some evidence that parasitization rates, and scale abundance, in the population as a whole were then more variable. But these effects were only slight and apparently short term, and there was certainly no evidence that scale abundance overall was increased by any diminution of the refuge effect.

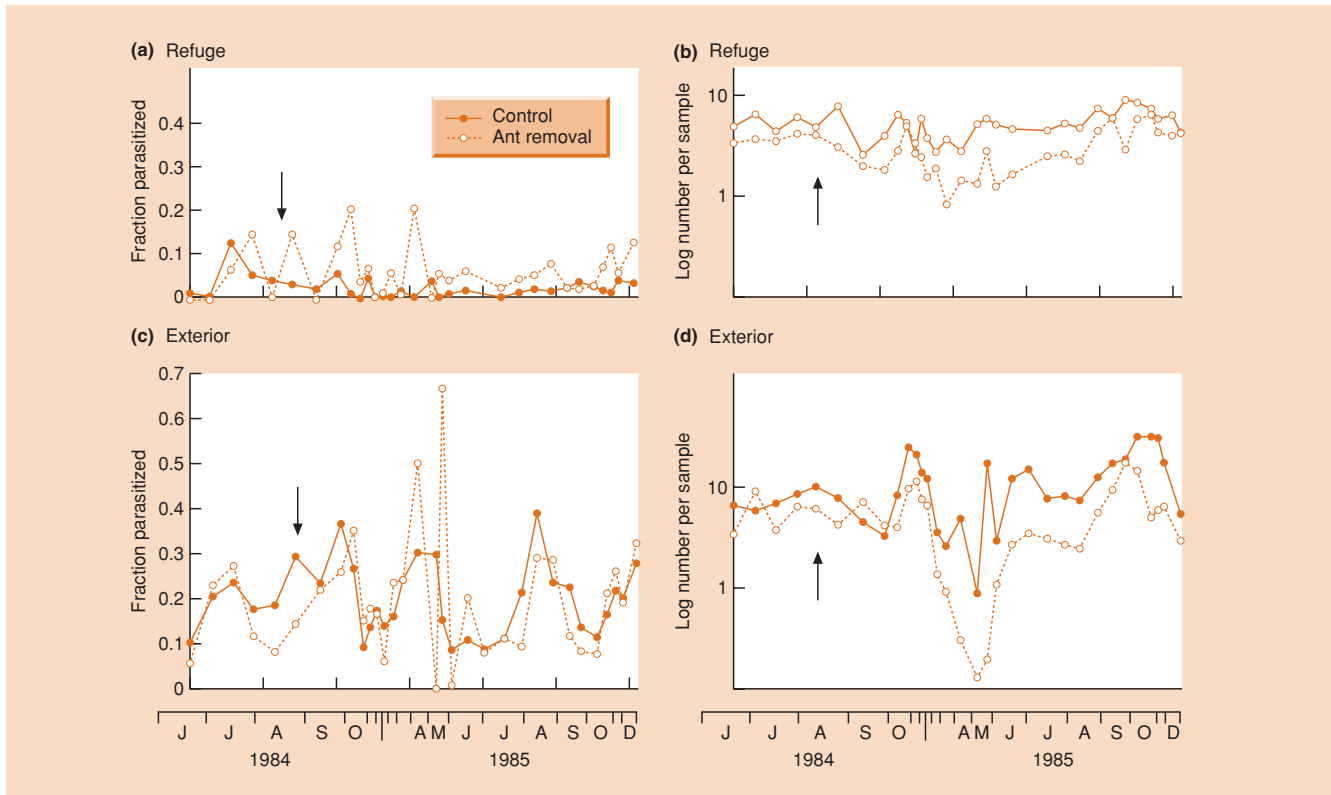


Figure 10.21 Results of a field experiment to test the hypothesis that the parasitoid *Aphytis mellitus* maintains the abundance of the California red scale, *Aonidiella aurantii*, at stable low levels because of a partial refuge from parasitization in the interior portions of citrus trees, where ants interfere with the parasitoids. When ants were removed from blocks of trees (time of removal indicated by the arrow), the fraction parasitized in the refuge tended to be higher (a), and scale abundance there was lower (b), but outside the refuge ('exterior') the fraction parasitized was only marginally more variable (c), and scale abundance was only more variable over one relatively brief period and tended to be lower than on control trees (d). (After Murdoch *et al.*, 1995.)

Moreover, Murdoch *et al.* (1985) had earlier argued that, in general, pest populations persist after successful biological control not as a result of aggregative responses, but because of the stochastic creation of host patches by colonization and their subsequent extinction when discovered by the agent: essentially, a metapopulation effect. Waage and Greathead (1988), however, suggested that a broader perspective could incorporate both aggregative responses and metapopulation effects. They proposed that scale insects and other homopterans, and mites (like Huffaker's), which may reproduce to have many generations within a patch, are often stabilized by asynchronies in the dynamics of different patches; whereas lepidopterans and hymenopterans, which typically occupy a patch for only part of a single generation, may often be stabilized by an aggregative response. In fact, though, with biological control, like predator-prey dynamics generally, building convincing links between patterns in population stability of natural populations and particular stabilizing mechanisms – or combinations of mechanisms – remains a challenge for the future.

10.6 Multiple equilibria: an explanation for outbreaks?

When predator and prey populations interact, there can sometimes be sudden changes in the abundance of one or both partners: outbreaks or crashes. Of course, this may reflect an equally sudden change in the environment, but ecologists working in a variety of fields have come to realize that there is not necessarily just one equilibrium combination of a predator and prey (about which there may or may not be oscillations). There can, instead, be 'multiple equilibria' or 'alternative stable states'.

Figure 10.22 is a model with multiple equilibria. The prey zero isocline has both a vertical section at low densities and a hump. This could reflect a type 3 functional response of a predator that also has a long handling time, or perhaps the combination of an aggregative response and an Allee effect in the prey. As a consequence, the predator zero

a model with multiple equilibria

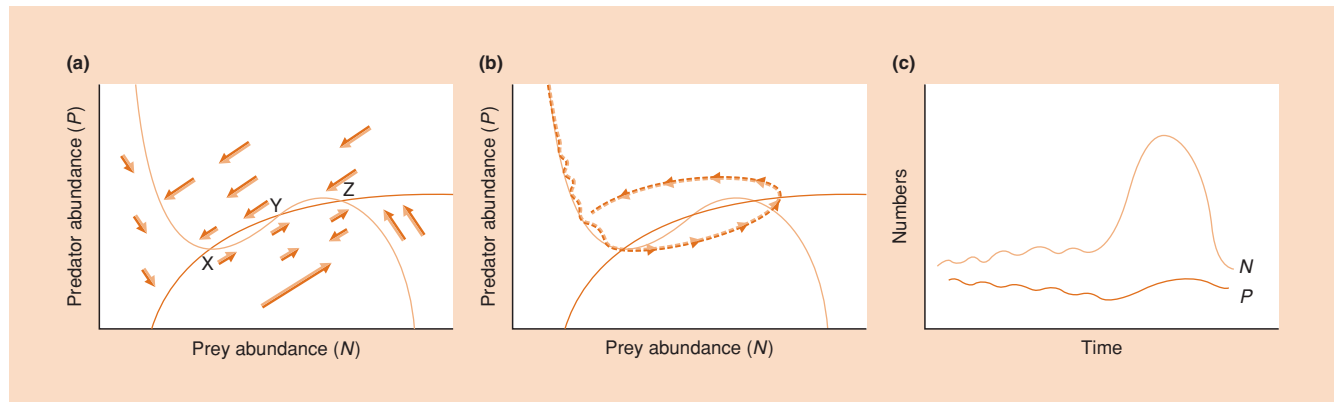


Figure 10.22 A predator–prey zero isocline model with multiple equilibria. (a) The prey zero isocline has a vertical section at low densities and a hump; the predator zero isocline can therefore cross it three times. Intersections X and Z are stable equilibria, but intersection Y is an unstable ‘breakpoint’ from which the joint abundances move towards either intersection X or intersection Z. (b) A feasible path that the joint abundances might take when subject to the forces shown in (a). (c) The same joint abundances plotted as numbers against time, showing that an interaction with characteristics that do not change can lead to apparent ‘outbreaks’ in abundance.

isocline crosses the prey zero isocline three times. The strengths and directions of the arrows in Figure 10.22a indicate that two of these points (X and Z) are fairly stable equilibria (although there are oscillations around each). The third point (Y), however, is unstable: populations near here will move towards either point X or point Z. Moreover, there are joint populations close to point X where the arrows lead to the zone around point Z, and joint populations close to point Z where the arrows lead back to the zone around point X. Even small environmental perturbations could put a population near point X on a path towards point Z, and vice versa.

The behavior of a hypothetical population, consistent with the arrows in Figure 10.22a, is plotted in Figure 10.22b on a joint abundance diagram, and in Figure 10.22c as a graph of numbers against time. The prey population, in particular, displays an ‘eruption’ in abundance, as it moves from a low-density equilibrium to a high-density equilibrium and back again. This eruption is in no sense a reflection of an equally marked change in the environment. It is, on the contrary, a pattern of abundance generated by the interaction itself (plus a small amount of environmental ‘noise’), and in particular it reflects the existence of multiple equilibria. Similar explanations may be invoked to explain apparently complicated patterns of abundance in nature.

There are certainly examples of natural populations exhibiting outbreaks of abundance from levels that are otherwise low and apparently stable (Figure 10.23a), and there are other examples in which populations appear to alternate between two stable densities (Figure 10.23b). But it does not follow that each of these examples is necessarily an interaction with multiple equilibria.

In some cases, a plausible argument for multiple equilibria can be put forward. This is true, for instance, of Clark’s (1964) work in Australia on the eucalyptus psyllid (*Cardiaspina albitextura*), a homopteran bug (Figure 10.23a).

These insects appear to have a low-density equilibrium maintained by their natural predators (especially birds), and a much less stable high-density equilibrium reflecting intraspecific competition (the destruction of host tree foliage leading to reductions in fecundity and survivorship). Outbreaks from one to the other can occur when there is just a short-term failure of the predators to react to an increase in the density of adult psyllids. Similarly, the observation of two alternative equilibria in Figure 10.23b for the viburnum whitefly, *Aleurotrachelus jelinekii*, is reinforced by a model for that population which predicts the same pattern (Southwood *et al.*, 1989).

Alternative stable states have also been proposed for a number of plant–herbivore interactions, often where increased grazing pressure seems to have led to the ‘collapse’ of the vegetation from a high biomass to a much lower one, which is then stable in the sense that there is no return to the high biomass state even when grazing pressure is severely reduced (van de Koppel *et al.*, 1997). The grasslands of the Sahel region of Africa, grazed by livestock, and the arctic plants along the coast of Hudson Bay in Canada, grazed by geese, are both examples. The conventional explanation (Noy-Meir, 1975) has essentially been that depicted in Figure 10.22: when driven to a low biomass, plants may have very little material above ground and hence very limited powers of immediate regrowth. This is a classic ‘Allee effect’ – the prey

sudden changes in abundance: multiple equilibria – or sudden changes in the environment

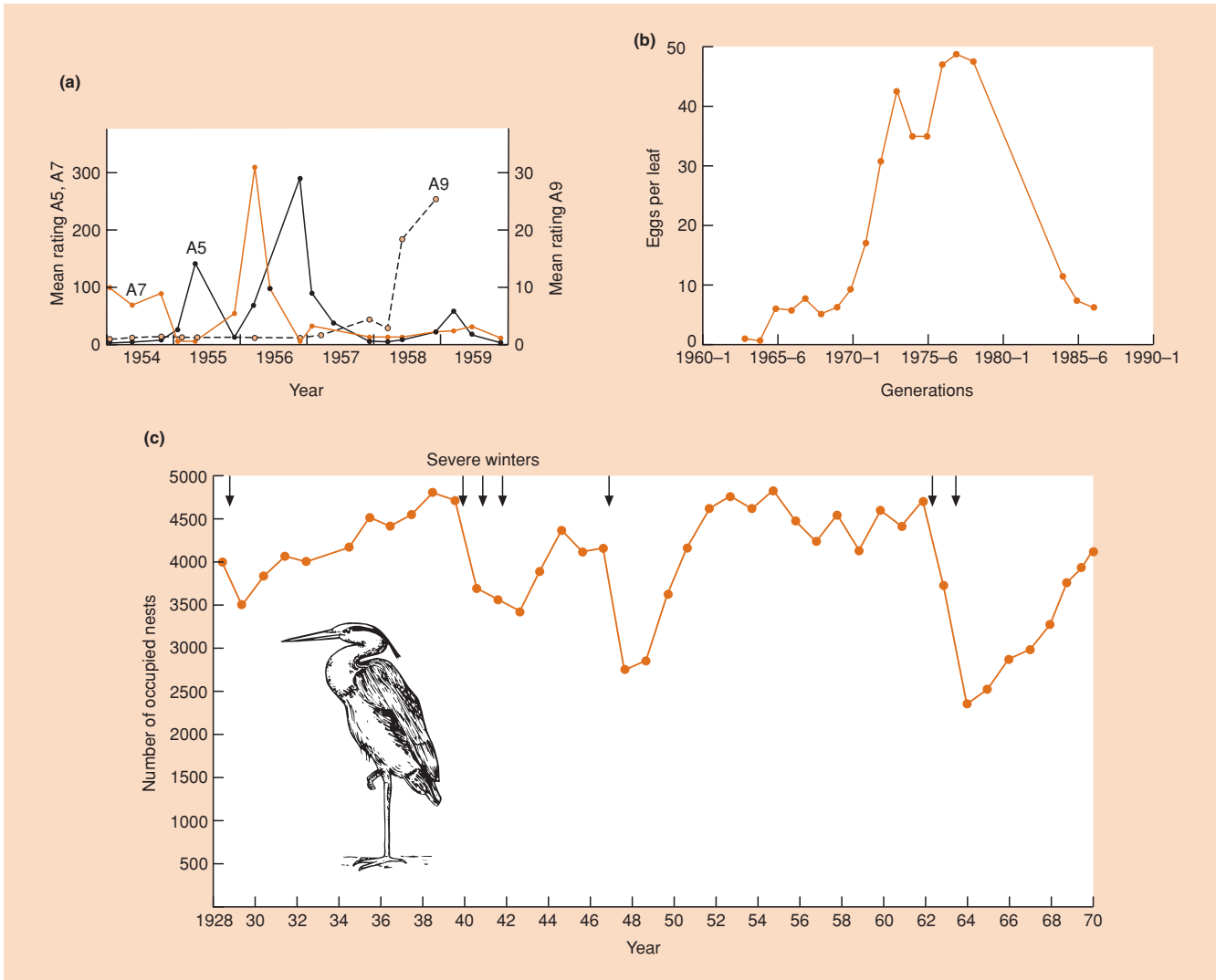


Figure 10.23 Possible examples of outbreaks and multiple equilibria. (a) Mean ratings of relative abundance of the eucalyptus psyllid, *Cardiaspina albitextura*, in three study areas in Australia (A5, A7 and A9). (After Clark, 1962.) (b) The mean number of eggs per leaf of the viburnum whitefly, *Aleurotrachelus jelinekii*, on a viburnum bush in Silwood Park, Berkshire, UK. No samples were taken between 1978 and 1979, and 1984 and 1985. (After Southwood *et al.*, 1989.) (c) Changes in the abundance of herons (*Ardea cinerea*) in England and Wales (measured by the number of nests occupied) are readily attributable to changes in environmental conditions (particularly severe winters). (After Stafford, 1971.)

suffering from too low an abundance – leading to a marked ‘hump’ in their isocline. It may also be, however, that the plants’ problems at low biomass are compounded by soil deterioration – erosion, for example – introducing further positive feedback into the system: high grazing leading to low plant biomass, leading to poorer growing conditions, leading to lower plant biomass, leading to even poorer growing conditions, and so on (van de Koppel *et al.*, 1997).

On the other hand, there are many cases in which sudden changes in abundance are fairly accurate reflections of sudden changes in the environment or a food source. For instance, the number of herons nesting in England and Wales normally fluctuates around 4000–4500 pairs, but the population declines markedly after particularly severe winters (Figure 10.23c). This fish-eating bird is unable to find sufficient food when inland waters become frozen for long periods, but there is no suggestion that the lower population levels (2000–3000 pairs) are an alternative equilibrium. The population crashes are simply the result of density-independent mortality from which the herons rapidly recover.

10.7 Beyond predator–prey

The simplest mathematical models of predator–prey interactions produce coupled oscillations that are highly unstable. However, by adding various elements of realism to these models it is possible to reveal the features of real predator–prey relationships that are likely to contribute to their stability. A further insight provided by models is that predator–prey systems may exist in more than one stable state. We have seen that a variety of patterns in the abundance of predators and prey, both in nature and in the laboratory, are consistent with the conclusions derived from models. Unfortunately, we are rarely in a position to apply specific explanations to particular sets of data, because the critical experiments and observations to test the models have rarely been made. Natural populations are affected not just by their predators or their prey, but also by many other environmental factors that serve to ‘muddy the waters’ when direct comparisons are made with simple models.

Moreover, the attention of both modelers and data gatherers (not that the two need be different) is increasingly being directed away from single- or two-species systems, towards those in which three species interact. For example, a pathogen attacking a predator that attacks a prey, or a parasitoid and a pathogen both attacking a prey/host. Interestingly, in several of these systems, unexpected dynamical properties emerge that are not just the expected blend of the component two-species interactions (Begon *et al.*, 1996; Holt, 1997). We return to the problems of ‘abundance’ in a broader context in Chapter 14.

Summary

Predator and prey populations display a variety of dynamic patterns. It is a major task for ecologists to account for the differences from one example to the next.

A number of mathematical models illustrate an underlying tendency for predator and prey populations to undergo coupled oscillations (cycles) in abundance. We explain the Lotka–Volterra model, which is the simplest differential equation predator–prey model, and using zero isoclines we show that the coupled oscillations are structurally unstable in this case. The model also illustrates the role of delayed density-dependent numerical responses in generating the cycles. We explain, too, the Nicholson–Bailey host–parasitoid model, which also displays unstable oscillations.

In both these models, cycles are several prey (host) generations in length, but other models of host–parasitoid (and host–pathogen) systems are able to generate coupled oscillations just one host generation in length.

We ask whether there is good evidence for predator–prey cycles in nature, focusing especially on a hare–lynx system and a moth attacked by two natural enemies. Even when predators or prey exhibit regular cycles in abundance, it is never easy to demonstrate that these are predator–prey cycles.

We begin an examination of the effects on dynamics of factors missing from the simplest models by looking at crowding. For predators, the most important expression of this is mutual interference. We look at the effects of crowding in the Lotka–Volterra model, including ratio-dependent predation: crowding stabilizes the dynamics, although this effect is strongest when the predators are least efficient. Essentially similar conclusions emerge from modifications of the Nicholson–Bailey model. There is, though, little direct evidence for these effects in nature.

The functional response describes the effect of prey abundance on predator consumption rate. The three types of functional response are explained, including the role of handling time in generating type 2 responses, and of variations in handling time and searching efficiency in generating type 3 responses. We explain the consequences for predator–prey dynamics of the different types of functional responses and of the ‘Allee effect’ (lowered recruitment at low abundance). Type 2 responses tend to destabilize, and type 3 responses to stabilize, but these are not necessarily important in practice.

Predators often (but not always) exhibit an aggregative response. We examine the effects of refuges and partial refuges in the Lotka–Volterra model, suggesting that spatial heterogeneities, and the responses to them, stabilize predator–prey dynamics, often at low prey densities. However, further work, especially with host–parasitoid systems and the Nicholson–Bailey model, shows that the effects of heterogeneity are complex. Stability arises through ‘aggregation of risk’, strengthening direct density dependencies that already exist. But aggregative responses that are spatially density dependent are least likely to

lead to aggregation of risk and least likely to enhance stability. Models with within-generation movement further undermine the significance of aggregative responses in stabilizing host-parasitoid interactions. A metapopulation perspective emphasizes that patch differences may stabilize through asynchrony, and also that predator-prey interactions may generate spatial as well as temporal patterns.

In practice, the stabilizing effects of metapopulation structure and of refuges have been demonstrated, and the general importance of responses to spatial heterogeneity in the choice of bio-control agents has been the subject of lively debate.

Finally, predator-prey systems with more than one equilibrium combination of predators and prey are examined as a possible basis for prey (or predator) outbreaks.



Chapter 11

Decomposers and Detritivores

11.1 Introduction

saprotrophs:
detritivores and
decomposers . . .

When plants and animals die, their bodies become resources for other organisms. Of course, in a sense, most consumers live on dead material – the carnivore catches and kills its prey, and the living leaf taken by a herbivore is dead by the time digestion starts. The critical distinction between the organisms in this chapter, and herbivores, carnivores and parasites, is that the latter all directly affect the rate at which their resources are produced. Whether it is lions eating gazelles, gazelles eating grass or grass parasitized by a rust fungus, the act of taking the resource harms the resource's ability to regenerate new resource (more gazelles or grass leaves). In contrast with these groups, saprotrophs (organisms that make use of dead organic matter) do not control the rate at which their resources are made available or regenerate; they are dependent on the rate at which some other force (senescence, illness, fighting, the shedding of leaves by trees) releases the resource on which they live. Exceptions exist among necrotrophic parasites (see Chapter 12) that kill and then continue to extract resources from the dead host. Thus, the fungus *Botrytis cinerea* attacks living bean leaves but continues this attack after the host's death. Similarly, maggots of the sheep blowfly *Lucilia cuprina* may parasitize and kill their host, whereupon they continue to feed on the corpse. In these cases the saprotroph can be said to have a measure of control over the supply of its food resource.

. . . do not generally
control their supply
of resources – 'donor
control'

We distinguish two groups of saprotrophs: decomposers (bacteria and fungi) and detritivores (animal consumers of dead matter). Pimm (1982) described the relationship that generally exists between decomposers or detritivores and their food as *donor controlled*: the donor (prey; i.e. dead organic matter) controls the density of the recipient (predator; i.e. decomposer or detritivore) but not the reverse. This

is fundamentally different from truly interactive predator-prey interactions (see Chapter 10). However, while there is generally no direct feedback between decomposers/detritivores and the dead matter consumed (and thus donor-controlled dynamics apply), nevertheless it is possible to see an indirect 'mutualistic' effect through the release of nutrients from decomposing litter, which may ultimately affect the rate at which trees produce more litter. In fact, it is in nutrient recycling that decomposers and detritivores play their most fundamental role (see Chapter 19). More generally, of course, the food webs associated with decomposition are just like food webs based on living plants: they have a number of trophic levels, including predators of decomposers (microbivores) and of detritivores, and consumers of these predators, and exhibit a range of trophic interactions (not just donor controlled).

Immobilization occurs when an inorganic nutrient element is incorporated into an organic form – primarily during the growth of green plants.

decomposition
defined

Conversely, decomposition involves the release of energy and the *mineralization* of chemical nutrients – the conversion of elements from an organic to inorganic form. Decomposition is defined as the gradual disintegration of dead organic matter and is brought about by both physical and biological agencies. It culminates with complex, energy-rich molecules being broken down by their consumers (decomposers and detritivores) into carbon dioxide, water and inorganic nutrients. Some of the chemical elements will have been locked up for a time as part of the body structure of the decomposer organisms, and the energy present in the organic matter will have been used to do work and is eventually lost as heat. Ultimately, the incorporation of solar energy in photosynthesis, and the immobilization of inorganic nutrients into biomass, is balanced by the loss of heat energy and organic nutrients when the organic matter is mineralized. Thus a given nutrient molecule may be successively immobilized and mineralized in a repeated round of nutrient cycling. We discuss the overall role played by decomposers and detritivores in the fluxes of energy

and nutrients at the ecosystem level in Chapters 17 and 18. In the present chapter, we introduce the organisms involved and look in detail at the ways in which they deal with their resources.

decomposition . . .
 . . . of dead
 bodies, . . .

It is not only the bodies of dead animals and plants that serve as resources for decomposers and detritivores. Dead organic matter is continually produced during the life of both animals and plants and can be a major resource. Unitary organisms shed dead parts as they develop and grow – the larval skins of arthropods, the skins of snakes, the skin, hair, feathers and horn of other vertebrates. Specialist feeders are often associated with these cast-off resources. Among the fungi there are specialist decomposers of feathers and of horn, and there are arthropods that specialize on sloughed off skin. Human skin is a resource for the household mites that are omnipresent inhabitants of house dust and cause problems for many allergy sufferers.

. . . of shed parts of
 organisms . . .

The continual shedding of dead parts is even more characteristic of modular organisms. Some polyps on a colonial hydroid or coral die and

decompose, while other parts of the same genet continue to regenerate new polyps. Most plants shed old leaves and grow new ones; the seasonal litter fall onto a forest floor is the most important of all the sources of resource for decomposers and detritivores, but the producers do not die in the process. Higher plants also continually slough off cells from the root caps, and root cortical cells die as a root grows through the soil. This supply of organic material from roots produces the very resource-rich *rhizosphere*. Plant tissues are generally leaky, and soluble sugars and nitrogenous compounds also become available on the surface of leaves, supporting the growth of bacteria and fungi in the *phyllosphere*.

. . . and of feces

Finally, animal feces, whether produced by detritivores, microbivores, herbivores, carnivores or parasites, are

a further category of resource for decomposers and detritivores. They are composed of dead organic material that is chemically related to what their producers have been eating.

The remainder of this chapter is in two parts. In Section 11.2 we describe the ‘actors’ in the saprotrophic ‘play’, and consider the relative roles of the bacteria and fungi on the one hand, and the detritivores on the other. Then, in Section 11.3, we consider, in turn, the problems and processes involved in the consumption by detritivores of plant detritus, feces and carrion.

11.2 The organisms

11.2.1 Decomposers: bacteria and fungi

If scavengers do not take a dead resource immediately it dies (such as hyenas consuming a dead zebra), the process of decomposition usually starts with colonization by bacteria and fungi. Other

changes may occur at the same time: enzymes in the dead tissue may start to autolyze it and break down the carbohydrates and proteins into simpler, soluble forms. The dead material may also become leached by rainfall or, in an aquatic environment, may lose minerals and soluble organic compounds as they are washed out in solution.

Bacteria and fungal spores are omnipresent in the air and the water, and are usually present on (and often in) dead material before it is dead.

bacteria and fungi
 are early colonists of
 newly dead material

They usually have first access to a resource. These early colonists tend to use soluble materials, mainly amino acids and sugars that are freely diffusible. They lack the array of enzymes necessary for digesting structural materials such as cellulose, lignin, chitin and keratin. Many species of *Penicillium*, *Mucor* and *Rhizopus*, the so-called ‘sugar fungi’ in soil, grow fast in the early phases of decomposition. Together with bacteria having similar opportunistic physiologies, they tend to undergo population explosions on newly dead substrates. As the freely available resources are consumed, these populations collapse, leaving very high densities of resting stages from which new population explosions may develop when another freshly dead resource becomes available. They may be thought of as the opportunist ‘r-selected species’ among the decomposers (see Section 4.12). Another example is provided by the early colonizers of nectar in flowers, predominantly yeasts (simple sugar fungi); these may spread to the ripe fruit where they act on sugar in the juice to produce alcohol (as happens in the industrial production of wine and beer).

In nature, as in industrial processes such as the making of wine or sauerkraut, the activity of the early colonizers is dominated by the metabolism of sugars and is strongly influenced by aeration. When oxygen is in free supply, sugars are metabolized to carbon dioxide by growing microbes. Under anaerobic conditions, fermentations produce a less complete breakdown of sugars to by-products such as alcohol and organic acids that change the nature of the environment for subsequent colonizers. In particular, the lowering of the pH by the production of acids has the effect of favoring fungal as opposed to bacterial activity.

domestic and
 industrial
 decomposition

Anoxic habitats are characteristic of waterlogged soils and, more particularly, of sediments of oceans and lakes. Aquatic sediments receive a continuous supply of dead organic matter from

aerobic and
 anaerobic
 decomposition
 in nature

the water column above but aerobic decomposition (mainly by bacteria) quickly exhausts the available oxygen because this can only be supplied from the surface of the sediment by diffusion. Thus, at some depth, from zero to a few centimeters below the surface, depending mainly on the load of organic material, sediments are completely anoxic. Below this level are found a variety of bacterial types that employ different forms of anaerobic respiration

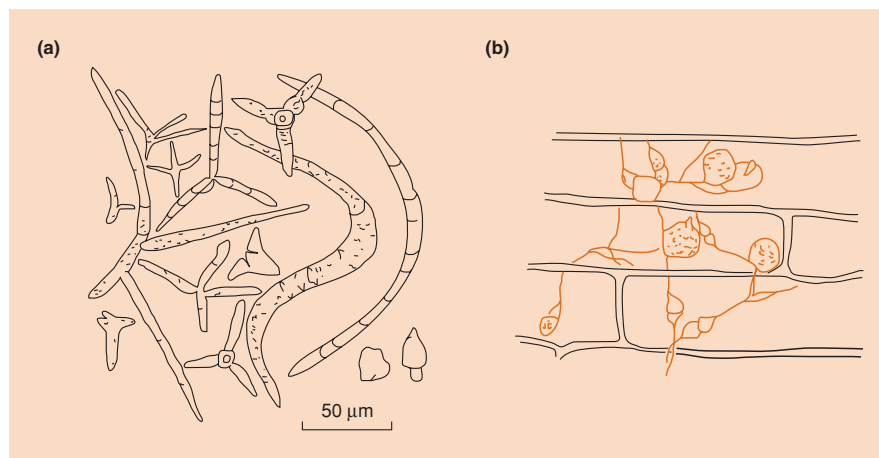


Figure 11.1 (a) Spores (conidia) of aquatic hyphomycete fungi from river foam. (b) Rhizomycelium of the aquatic fungus *Cladochytrium replicatum* within the epidermis of an aquatic plant. The circular bodies are zoosporangia. (After Webster, 1970.)

– that is, they use terminal inorganic electron acceptors other than oxygen in their respiratory process. The bacterial types occur in a predictable pattern with denitrifying bacteria at the top, sulfate-reducing bacteria next and methanogenic bacteria in the deepest zone. Sulfate is comparatively abundant in sea water and so the zone of sulfate-reducing bacteria is particularly wide (Fenchel, 1987b). In contrast, the concentration of sulfate in lakes is low, and methanogenesis plays a correspondingly larger role (Holmer & Storkholm, 2001).

A strong element of chance determines which species are the first to colonize newly dead material, but in some environments there are specialists with properties that enhance their chances of arriving early. Litter that falls into streams or ponds is often colonized by aquatic fungi (e.g. Hyphomycetes), which bear spores with sticky tips (Figure 11.1a) and are often of a curious form that seems to maximize their chance of being carried to and sticking to leaf litter. They may spread by growing from cell to cell within the tissues (Figure 11.1b).

decomposition of more resistant tissues proceeds more slowly

After the colonization of terrestrial litter by the ‘sugar’ fungi and bacteria, and perhaps also after leaching by rain or in the water, the residual resources are not diffusible and are more resistant

to attack. In broad terms, the major components of dead terrestrial organic matter are, in a sequence of increasing resistance to decomposition: sugars < (less resistant than) starch < hemicelluloses, pectins and proteins < cellulose < lignins < suberins < cutins. Hence, after an initial rapid breakdown of sugar, decomposition proceeds more slowly, and involves microbial specialists that can use celluloses and lignins and break down the more complex proteins, suberin (cork) and cuticles. These are structural compounds, and their breakdown and metabolism depend on very intimate contact with the decomposers (most cellulases are surface enzymes requiring actual physical contact between the decomposer organism and its resource). The processes of

decomposition may now depend on the rate at which fungal hyphae can penetrate from cell to cell through lignified cell walls. In the decomposition of wood by fungi (mainly homobasidiomycetes), two major categories of specialist decomposers can be recognized: the brown rots that can decompose cellulose but leave a predominantly lignin-based brown residue, and the white rots that decompose mainly the lignin and leave a white cellulosic residue (Worrall *et al.*, 1997). The tough silicon-rich frustules of dead diatoms in the phytoplankton communities of lakes and oceans are somewhat analogous to the wood of terrestrial communities. The regeneration of this silicon is critical for new diatom growth, and decomposition of the frustules is brought about by specialized bacteria (Bidle & Azam, 2001).

The organisms capable of dealing with progressively more refractory compounds in terrestrial litter represent a natural succession starting with simple sugar fungi (mainly Phycocomycetes and Fungi Imperfecti), usually followed by septate fungi (Basidiomycetes and Actinomycetes) and Ascomycetes, which are slower growing, spore less freely, make intimate contact with their substrate and have more specialized metabolism. The diversity of the microflora that decomposes a fallen leaf tends to decrease as fewer but more highly specialized species are concerned with the last and most resistant remains.

succession of decomposing microorganisms

The changing nature of a resource during its decomposition is illustrated in Figure 11.2a for beech leaf litter on the floor of a cool temperate deciduous forest in Japan. Polyphenols and soluble carbohydrates quickly disappeared, but the resistant structural holocellulose and lignin decomposed much more slowly. The fungi responsible for leaf decomposition follow a succession that is associated with the changing nature of the resource. The frequency of occurrence of early species, such as *Arthrimum* sp. (Figure 11.2b), was correlated with declines in holocellulose and soluble carbohydrate concentrations; Osono and Takeda (2001) suggest that they

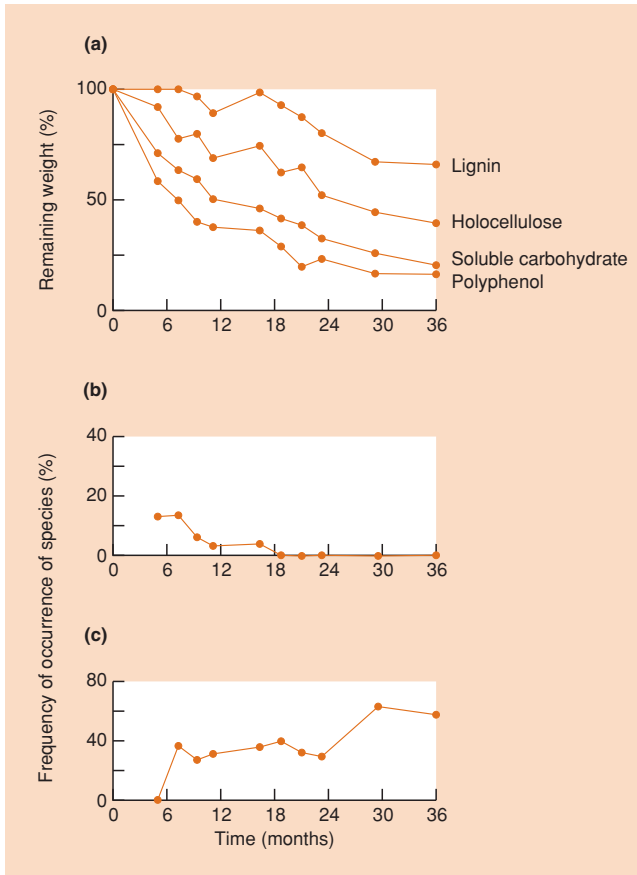


Figure 11.2 (a) Changes in the composition of beech (*Fagus crenata*) leaf litter (in mesh bags) during decomposition on a woodland floor in Japan over a 3-year period. Amounts are expressed as percentages of the starting quantities. (b, c) Changes in the frequency of occurrence of fungal species representative of: (b) early species (*Arthrimum* sp.) and (c) late species (*Mortierella ramanniana*). (After Osono & Takeda, 2001.)

depend on these components for their growth. Many late species, such as *Mortierella ramanniana*, seem to rely on sugars released by other fungi capable of decomposing lignin.

most microbial decomposers are relatively specialized

Individual species of microbial decomposer are not biochemically very versatile; most of them can cope with only a limited number of substrates. It is the diversity of species

involved that allows the structurally and chemically complex tissues of a plant or animal corpse to be decomposed. Between them, a varied microbiota of bacteria and fungi can accomplish the complete degradation of dead material of both plants and animals. However, in practice they seldom act alone, and the process would be much slower and, moreover, incomplete, if they did so. The major factor that delays the decomposition of

organic residues is the resistance to decomposition of plant cell walls – an invading decomposer meets far fewer barriers in an animal body. The process of plant decomposition is enormously speeded up by any activity that grinds up and fragments the tissues, such as the chewing action of detritivores. This breaks open cells and exposes the contents and the surfaces of cell walls to attack.

11.2.2 Detritivores and specialist microbivores

The microbivores are a group of animals that operate alongside the detritivores, and which can be difficult to distinguish from them. The name microbivore is reserved for the minute animals that

specialist consumers of microbial organisms: microbivores

specialize at feeding on microflora, and are able to ingest bacteria or fungi but exclude detritus from their guts. Exploitation of the two major groups of microflora requires quite different feeding techniques, principally because of differences in growth form. Bacteria (and yeasts) show a colonial growth form arising by the division of unicells, usually on the surface of small particles. Specialist consumers of bacteria are inevitably very small; they include free-living protozoans such as amoebae, in both soil and aquatic environments, and the terrestrial nematode *Pelodera*, which does not consume whole sediment particles but grazes among them consuming the bacteria on their surfaces. The majority of fungi, in contrast to most bacteria, are filamentous, producing extensively branching hyphae, which in many species are capable of penetrating organic matter. Some specialist consumers of fungi possess piercing, sucking stylets (e.g. the nematode *Ditylenchus*) that they insert into individual fungal hyphae. However, most fungivorous animals graze on the hyphae and consume them whole. In some cases, close mutualistic relationships exist between fungivorous beetles, ants and termites and characteristic species of fungi. These mutualisms are discussed in Chapter 13.

Note that microbivores consume a living resource and may not be subject to donor-controlled dynamics (Laakso *et al.*, 2000). In a study of decomposition of lake weed and phytoplankton in laboratory microcosms, Jurgens and Sala (2000) followed the fate of bacteria (decomposers) in the presence and absence of bacteria-grazing protists, namely *Spumella* sp. and *Bodo saltans* (microbivores). In the presence of the microbivores, there was a reduction of 50–90% in bacterial biomass and the bacterial community became dominated by large, grazer-resistant forms including filamentous bacteria.

The larger the animal, the less able it is to distinguish between microflora as food and the plant or animal detritus on which these are growing. In fact, the majority of the detritivorous animals involved in the decomposition of dead organic matter are generalist consumers, of both the detritus itself and the associated microfloral populations.

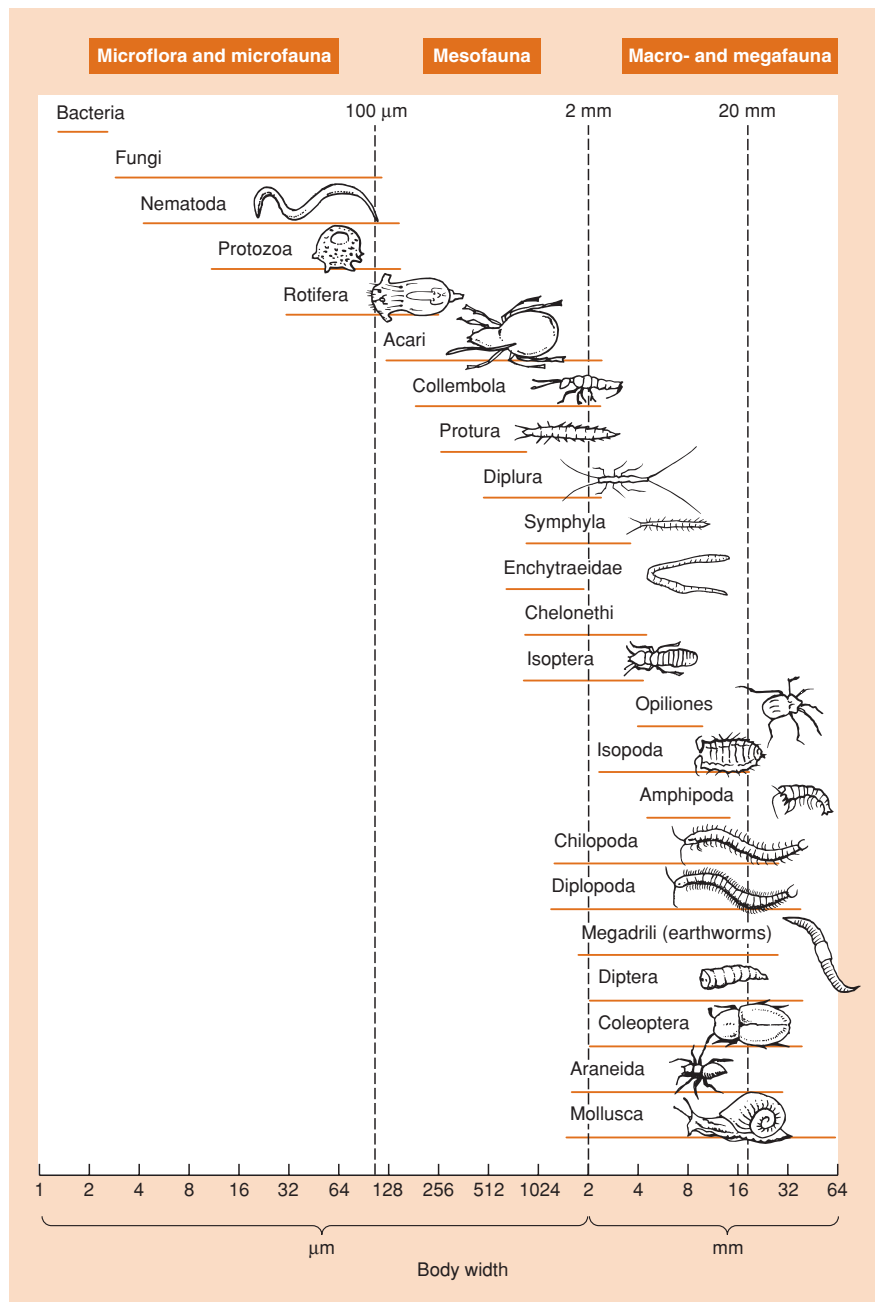


Figure 11.3 Size classification by body width of organisms in terrestrial decomposer food webs. The following groups are wholly carnivorous: Opiliones (harvest spiders), Chilopoda (centipedes) and Araneida (spiders). (After Swift *et al.*, 1979.)

classification of decomposers . . .
 . . . by size in terrestrial environments . . .

The protists and invertebrates that take part in the decomposition of dead plant and animal materials are a taxonomically diverse group. In terrestrial environments they are usually classified according to their size. This is not an arbitrary basis for classification, because size is an important feature for organisms that reach their resources by burrowing or crawling among cracks and crevices of litter or soil. The

microfauna (including the specialist microbivores) includes protozoans, nematode worms and rotifers (Figure 11.3). The principal groups of the *mesofauna* (animals with a body width between 100 μm and 2 mm) are litter mites (Acari), springtails (Collembola) and pot worms (Enchytraeidae). The *macrofauna* (2–20 mm body width) and, lastly, the *megafauna* (> 20 mm) include woodlice (Isopoda), millipedes (Diplopoda), earthworms (Megadrili), snails and slugs (Mollusca) and the larvae of certain flies (Diptera) and beetles (Coleoptera). These animals are mainly responsible for the

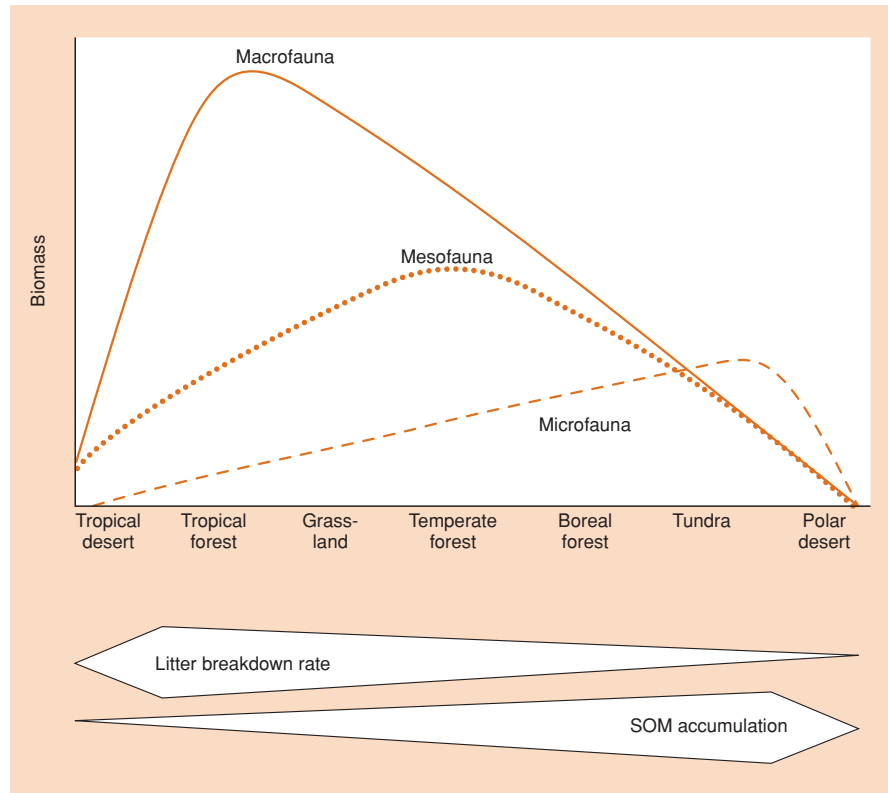


Figure 11.4 Patterns of latitudinal variation in the contribution of the macro-, meso- and microfauna to decomposition in terrestrial ecosystems. Soil organic matter (SOM) accumulation (inversely related to litter breakdown rate) is promoted by low temperatures and waterlogging, where microbial activity is impaired. (Swift *et al.*, 1979.)

initial shredding of plant remains. By their action, they may bring about a large-scale redistribution of detritus and thus contribute directly to the development of soil structure. It is important to note that the microfauna, with their short generation times, operate at the same scale as bacteria and can track bacterial population dynamics, whilst the mesofauna and the fungi they mainly depend on are both longer lived. The largest and longest lived detritivores, in contrast, cannot be finely selective in their diet, but choose patches of high decomposer activity (J. M. Anderson, personal communication).

Long ago, Charles Darwin (1888) estimated that earthworms in some pastures close to his house formed a new layer of soil 18 cm deep in 30 years, bringing about 50 tons ha^{-1} to the soil surface each year as worm casts. Figures of this order of magnitude have since been confirmed on a number of occasions. Moreover, not all species of earthworm put their casts above ground, so the total amount of soil and organic matter that they move may be much greater than this. Where earthworms are abundant, they bury litter, mix it with the soil (and so expose it to other decomposers and detritivores), create burrows (so increasing soil aeration and drainage) and deposit feces rich in organic matter. It is not surprising that agricultural ecologists become worried about practices that reduce worm populations.

Detritivores occur in all types of terrestrial habitat and are often found at remarkable species richness and in very great numbers.

Thus, for example, a square meter of temperate woodland soil may contain 1000 species of animals, in populations exceeding 10 million for nematode worms and protozoans, 100,000 for springtails (*Collembola*) and soil mites (*Acari*), and 50,000 or so for other invertebrates (Anderson, 1978). The relative importance of microfauna, mesofauna and macrofauna in terrestrial communities varies along a latitudinal gradient (Figure 11.4). The microfauna is relatively more important in the organic soils in boreal forest, tundra and polar desert. Here the plentiful organic matter stabilizes the moisture regime in the soil and provides suitable microhabitats for the protozoans, nematodes and rotifers that live in interstitial water films. The hot, dry, mineral soils of the tropics have few of these animals. The deep organic soils of temperate forests are intermediate in character; they maintain the highest mesofaunal populations of litter mites, springtails and pot worms. The majority of the other soil animal groups decline in numbers towards the drier tropics, where they are replaced by termites. Lower mesofaunal diversity in these tropical regions may be related to a lack of litter due to decomposition and consumption by termites, reflecting both low resource abundance and few available microhabitats (J. M. Anderson, personal communication).

On a more local scale, too, the nature and activity of the decomposer community depends on the conditions in which the organisms live. Temperature has a fundamental role in determining

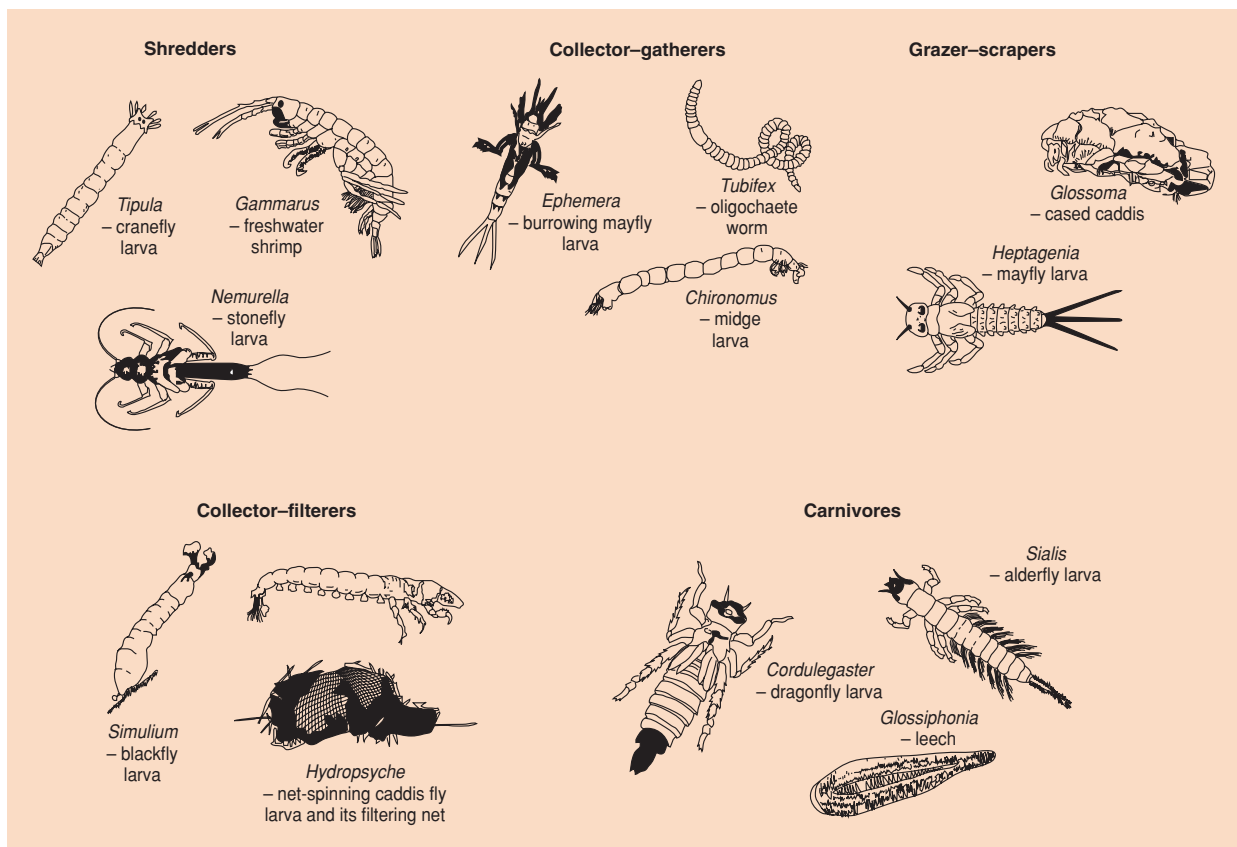


Figure 11.5 Examples of the various categories of invertebrate consumer in freshwater environments.

the rate of decomposition and, moreover, the thickness of water films on decomposing material places absolute limits on mobile microfauna and microflora (protozoa, nematode worms, rotifers and those fungi that have motile stages in their life cycles). In dry soils, such organisms are virtually absent. A continuum can be recognized from dry conditions through waterlogged soils to true aquatic environments. In the former, the amount of water and thickness of water films are of paramount importance, but as we move along the continuum, conditions change to resemble more and more closely those of the bed of an open-water community, where oxygen shortage, rather than water availability, may dominate the lives of the organisms.

In freshwater ecology the study of detritivores has been concerned less with the size of the organisms than with the ways in which they obtain their food. Cummins (1974) devised a scheme that recognizes four main categories of invertebrate consumer in streams. *Shredders* are detritivores that feed on coarse particulate organic matter (particles > 2 mm in size), and during feeding these serve to fragment the material. Very often

... and by feeding
mode in aquatic
environments

in streams, the shredders, such as cased caddis-fly larvae of *Stenophylax* spp., freshwater shrimps (*Gammarus* spp.) and isopods (e.g. *Asellus* spp.), feed on tree leaves that fall into the stream. *Collectors* feed on fine particulate organic matter (< 2 mm). Two subcategories of collectors are defined. *Collector-gatherers* obtain dead organic particles from the debris and sediments on the bed of the stream, whereas *collector-filterers* sift small particles from the flowing column of water. Some examples are shown in Figure 11.5. *Grazer-scrappers* have mouthparts appropriate for scraping off and consuming the organic layer attached to rocks and stones; this organic layer is comprised of attached algae, bacteria, fungi and dead organic matter adsorbed to the substrate surface. The final invertebrate category is *carnivores*. Figure 11.6 shows the relationships amongst these invertebrate feeding groups and three categories of dead organic matter. This scheme, developed for stream communities, has obvious parallels in terrestrial ecosystems (Anderson, 1987) as well as in other aquatic ecosystems. Earthworms are important shredders in soils, while a variety of crustaceans perform the same role on the sea bed. On the other hand, filtering is common among marine but not terrestrial organisms.

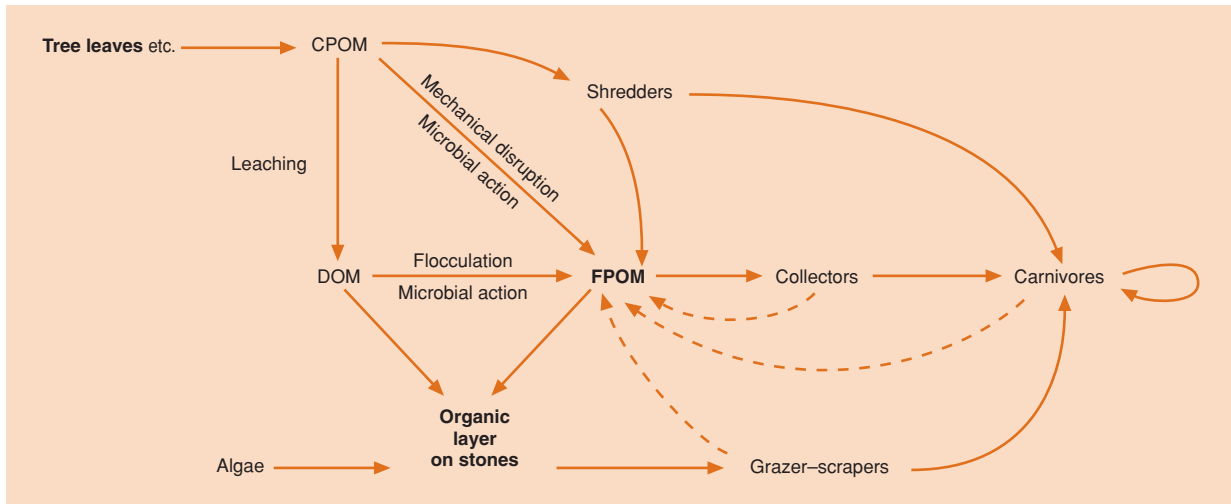


Figure 11.6 A general model of energy flow in a stream. A fraction of coarse particulate organic matter (CPOM) is quickly lost to the dissolved organic matter (DOM) compartment by leaching. The remainder is converted by three processes to fine particulate organic matter (FPOM): (i) mechanical disruption by battering; (ii) processing by microorganisms causing gradual break up; and (iii) fragmentation by the shredders. Note also that all animal groups contribute to FPOM by producing feces (dashed lines). DOM is also converted into FPOM by a physical process of flocculation or via uptake by microorganisms. The organic layer attached to stones on the stream bed derives from algae, DOM and FPOM adsorbed onto an organic matrix.

The feces and bodies of aquatic invertebrates are generally processed along with dead organic matter from other sources by shredders and collectors. Even the large feces of aquatic vertebrates do not appear to possess a characteristic fauna, probably because such feces are likely to fragment and disperse quickly as a result of water movement. Carrion also lacks a specialized fauna – many aquatic invertebrates are omnivorous, feeding for much of the time on plant detritus and feces with their associated microorganisms, but ever ready to tackle a piece of dead invertebrate or fish when this is available. This contrasts with the situation in the terrestrial environment, where both feces and carrion have specialized detritivore faunas (see Sections 11.3.3 and 11.3.5).

detritivore-dominated communities

Some animal communities are composed almost exclusively of detritivores and their predators. This is true not only of the forest floor, but also of shaded streams, the depths of oceans and lakes, and the permanent residents of caves: in short, wherever there is insufficient light for appreciable photosynthesis but nevertheless an input of organic matter from nearby plant communities. The forest floor and shaded streams receive most of their organic matter as dead leaves from trees. The beds of oceans and lakes are subject to a continuous settlement of detritus from above. Caves receive dissolved and particulate organic matter percolating down through soil and rock, together with windblown material and the debris of migrating animals.

11.2.3 The relative roles of decomposers and detritivores

The roles of the decomposers and detritivores in decomposing dead organic matter can be compared in a variety of ways. A comparison of numbers will reveal a predominance of bacteria. This is almost inevitable because we are counting individual cells. A comparison of biomass gives a quite different picture. Figure 11.7 shows the relative amounts of biomass represented in different groups involved in the decomposition of litter on a forest floor (expressed as the relative amounts of nitrogen present). For most of the year, decomposers (microorganisms) accounted for five to 10 times as much of the biomass as the detritivores. The biomass of detritivores varied less through the year because they are less sensitive to climatic change, and they were actually predominant during a period in the winter.

Unfortunately, the biomass present in different groups of decomposers is itself a poor measure of their relative importance in the process of decomposition. Populations of organisms with short lives and high activity may contribute more to the activities in the community than larger, long-lived, sluggish species (e.g. slugs!) that make a greater contribution to biomass.

Lillebo *et al.* (1999) attempted to distinguish the relative roles, in the

assessing the relative importance of decomposers and detritivores . . .

. . . in the decomposition of a salt marsh plant, . . .

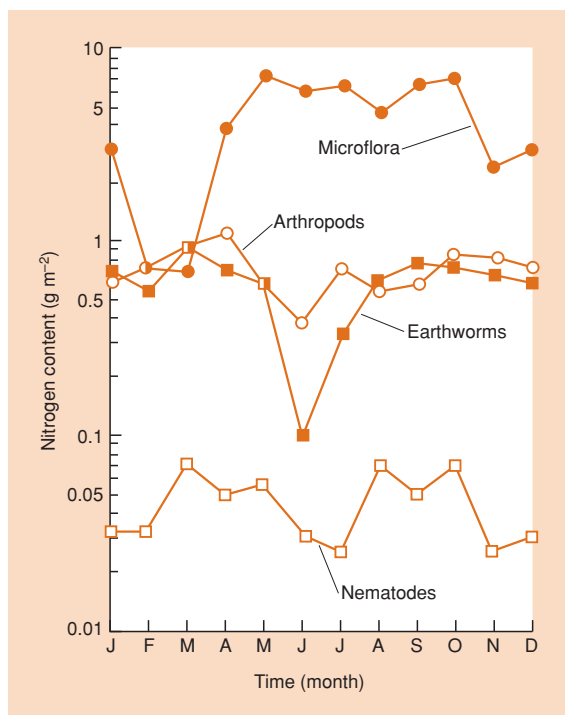


Figure 11.7 The relative importance in forest litter decomposition of microflora in comparison with arthropods, earthworms and nematodes, expressed in terms of their relative content of nitrogen – a measure of their biomass. Microbial activity is much greater than that of detritivores but the latter is more constant through the year. (After Ausmus *et al.*, 1976.)

decomposition of the salt marsh plant *Spartina maritima*, of bacteria, microfauna (e.g. flagellates) and macrofauna (e.g. the snail *Hydrobia ulvae*) by creating artificial communities in laboratory microcosms. At the end of the 99-day study, 32% of the biomass

of *Spartina* leaves remained in the bacteria treatment, whereas only 8% remained when the microfauna and macrofauna were also present (Figure 11.8a). Separate analyses of the mineralization of the carbon, nitrogen and phosphorus content of the leaves also revealed that bacteria were responsible for the majority of the mineralization, but that microfauna and particularly macrofauna enhanced the mineralization rates in the case of carbon and nitrogen (Figure 11.8b).

The decomposition of dead material is not simply due to the sum of the activities of microbes and detritivores: it is largely the result of interaction between the two. The shredding action of detritivores, such as the snail *Hydrobia ulvae* in the experiment of Lillebo *et al.* (1999), usually produces smaller particles with a larger surface area (per unit volume of litter) and thus increases the area of substrate available for microorganism growth. In addition, the activity of fungi may be stimulated by the disruption, through grazing, of competing hyphal networks. Moreover, the activity of both fungi and bacteria may be enhanced by the addition of mineral nutrients in urine and feces (Lussenhop, 1992).

The ways in which the decomposers and detritivores interact might be studied by following a leaf fragment through the process of decomposition, focusing attention on a part of the wall of a single cell. Initially, when the leaf falls to the ground, the piece of cell wall is protected from microbial attack because it lies within the plant tissue. The leaf is now chewed and the fragment enters the gut of, say, an isopod. Here it meets a new microbial flora in the gut and is acted on by the digestive enzymes of the isopod. The fragment emerges, changed by its passage through the gut. It is now part of the isopod's feces and is much more easily attacked by microorganisms, because it has been fragmented and partially digested. While microorganisms are colonizing, it may again be

... in a terrestrial environment, ...

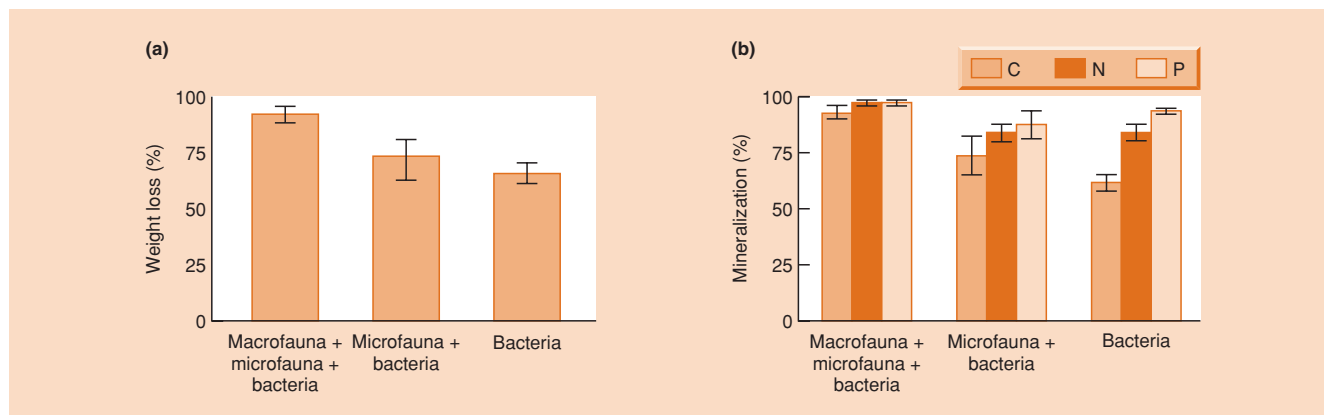


Figure 11.8 (a) Weight loss of *Spartina maritima* leaves during 99 days in the presence of: (i) macrofauna + microfauna + bacteria, (ii) microfauna + bacteria, or (iii) bacteria alone (mean \pm SD). (b) Percentage of initial carbon, nitrogen and phosphorus content that was mineralized during 99 days in the three treatments. (After Lillebo *et al.*, 1999.)

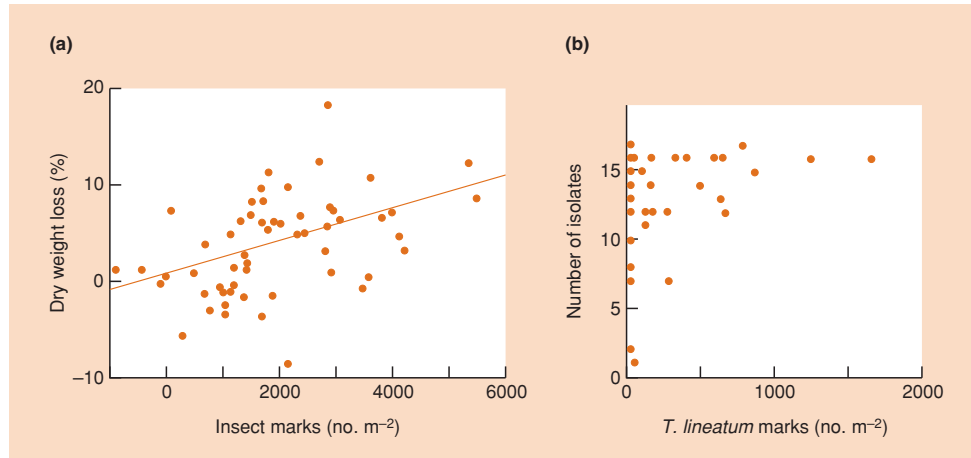


Figure 11.9 Relationships between (a) the decay of standard pieces of dead spruce wood over a 2.5-year period in Finland and the number of insect marks, and (b) the fungal infection rate (number of fungal isolates per standard piece of wood) and number of marks made by the beetle *Tripodendron lineatum*. Dry weight loss and number of insect marks in (a) were obtained by subtracting the values for each wood sample held in a permanently closed net cage from the corresponding value for its counterpart in a control cage that permitted insect entry. In some cases, the dry weight loss of the counterpart wood sample was lower, so the percentage weight loss was negative. This is possible because the number of insect visits does not explain all the variation in dry weight loss. (After Muller *et al.*, 2002.)

eaten, perhaps by a coprophagous springtail, and pass through the new environment of the springtail's gut. Incompletely digested fragments may again appear, this time in springtail feces, yet more easily accessible to microorganisms. The fragment may pass through several other guts in its progress from being a piece of dead tissue to its inevitable fate of becoming carbon dioxide and minerals.

... in a freshwater environment, ...

Fragmentation by detritivores plays a key role in terrestrial situations because of the tough cell walls characteristic of vascular plant detritus. The same is true in many freshwater environments where terrestrial litter makes up most of the available detritus. In contrast, detritus at the lowest trophic level in marine environments consists of phytoplankton cells and seaweeds; the former present a high surface area without the need for physical disruption and the latter, lacking the structural polymers of vascular plant cell walls, are prone to fragmentation by physical factors. Rapid decomposition of marine detritus is probably less dependent on fragmentation by invertebrates; shredders are rare in the marine environment compared to its terrestrial and freshwater counterparts (Plante *et al.*, 1990).

... in dead wood ...

Dead wood provides particular challenges to colonization by microorganisms because of its patchy distribution and tough exterior. Insects can enhance fungal colonization of dead wood by carrying fungi to their 'target' or by enhancing access of air-disseminated fungal propagules by making holes in the outer bark into the phloem and xylem. Muller *et al.* (2002)

distributed standard pieces of spruce wood (*Picea abies*) on a forest floor in Finland. After 2.5 years, the numbers of insect 'marks' (boring and gnawing) were recorded and were found to be correlated with dry weight loss of the wood (Figure 11.9a). This relationship comes about because of biomass consumption by the insects but also, to an unknown extent, by fungal action that has been enhanced by insect activity. Thus, fungal infection rate was always high when there were more than 400 marks per piece of wood made by the common ambrosia beetle *Tripodendron lineatum* (Figure 11.9b). This species burrows deeply into the sapwood and produces galleries about 1 mm in diameter. Some of the fungal species involved are likely to have been transmitted by the beetle (e.g. *Ceratocystis piceae*) but the invasion of other, air-disseminated types is likely to have been promoted by the galleries left by the beetle.

The enhancement of microbial respiration by the action of detritivores has also been reported in the decomposition of small mammal carcasses.

... and in small mammal carcasses

Two sets of insect-free rodent carcasses weighing 25 g were exposed under experimental conditions in an English grassland in the fall. In one set the carcasses were left intact. In the other, the bodies were artificially riddled with tunnels by repeated piercing of the material with a dissecting needle to simulate the action of blowfly larvae in the carcass. The results of this experiment paralleled those of the wood decomposition study above; here, the tunnels enhanced microbial activity (Figure 11.10) by disseminating the microflora as well as increasing the aeration of the carcass.

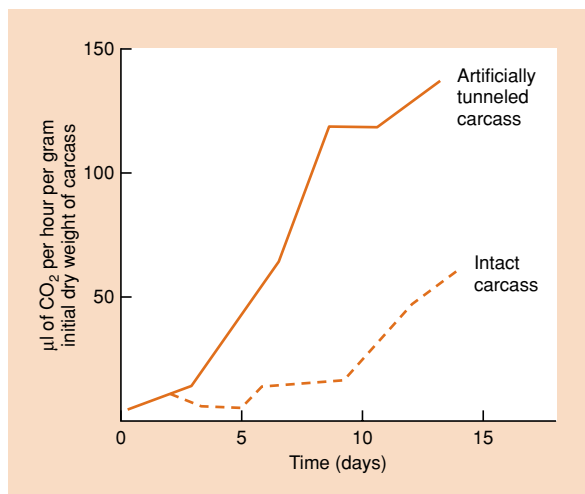


Figure 11.10 The evolution of carbon dioxide (CO_2), a measure of microbial activity, from carcasses of small mammals placed in ‘respiration’ cylinders and screened from insect attack. One set of carcasses was left intact, while the second set was pierced repeatedly with a dissecting needle to simulate the action of tunneling by blowfly larvae. (After Putman, 1978a.)

11.2.4 Ecological stoichiometry and the chemical composition of decomposers, detritivores and their resources

‘ecological stoichiometry’ and relations between resources and consumers

Ecological stoichiometry, defined by Elser and Urabe (1999) as the analysis of constraints and consequences in ecological interactions of the mass balance of multiple chemical elements (particularly the ratios of carbon to nitrogen and of carbon to phosphorus),

is an approach that can shed light on the relations between resources and consumers. Many studies have focused on plant–herbivore relations (Hessen, 1997) but the approach is also important when considering decomposers, detritivores and their resources.

There is a great contrast between the chemical composition of dead plant tissue and that of the tissues of the heterotrophic organisms that consume and decompose it. While the major components of plant tissues, particularly cell walls, are structural polysaccharides, these are only of minor significance in the bodies of microorganisms and detritivores. However, being harder to digest than storage carbohydrates and protein, the structural chemicals still form a significant component of detritivore feces. Detritivore feces and plant tissue have much in common chemically, but the protein and lipid contents of detritivores and decomposers are significantly higher than those of plants and feces.

The rate at which dead organic matter decomposes is strongly dependent on its biochemical composition. This is because microbial tissue has very high nitrogen and phosphorus contents, indicative of high requirements for these nutrients. Roughly speaking, the stoichiometric ratios of carbon : nitrogen (C : N) and carbon : phosphorus (C : P) in decomposers are 10 : 1 and 100 : 1, respectively (e.g. Goldman *et al.*, 1987). In other words, a microbial population of 111 g can only develop if there is 10 g of nitrogen and 1 g of phosphorus available. Terrestrial plant material has much higher ratios, ranging from 19 to 315 : 1 for C : N and from 700 to 7000 : 1 for C : P (Enriquez *et al.*, 1993). Consequently, this material can support only a limited biomass of decomposer organisms and the whole pace of the decomposition process will itself be limited by nutrient availability. Marine and freshwater plants and algae tend to have ratios more similar to the decomposers (Duarte, 1992), and their rates of decomposition are correspondingly faster (Figure 11.11a). Figure 11.11b and c illustrate the strong relationships between initial nitrogen and phosphorus concentration in plant tissue and its decomposition rate for a wide range of plant detritus from terrestrial, freshwater and marine species.

decomposition rate depends on ...
... biochemical composition ...

The rate at which dead organic matter decomposes is also influenced by inorganic nutrients, especially nitrogen (as ammonium or nitrate), that are available from the environment. Thus, greater microbial biomass can be supported, and decomposition proceeds faster, if nitrogen is absorbed from outside. For example, grass litter decomposes faster in streams running through tussock grassland in New Zealand that has been improved for pasture (where the water is, in consequence, richer in nitrate) than in ‘unimproved’ settings (Young *et al.*, 1994).

... and mineral nutrients in the environment

One consequence of the capacity of decomposers to use inorganic nutrients is that after plant material is added to soil, the level of soil nitrogen tends to fall rapidly as it is incorporated into microbial biomass.

The effect is particularly evident in agriculture, where the ploughing in of stubble can result in nitrogen deficiency of the subsequent crop. In other words, the decomposers compete with the plants for inorganic nitrogen. This raises a significant and somewhat paradoxical issue. We have noted that plants and decomposers are linked by an indirect *mutualism* mediated by nutrient recycling – plants provide energy and nutrients in organic form that are used by decomposers, and decomposers mineralize the organic material back to an inorganic form that can again be used by plants. However, stoichiometric constraints on carbon and nutrients also lead to *competition* between the plants and decomposers (usually for nitrogen in terrestrial communities, often

complex relationships between decomposers and living plants, ...

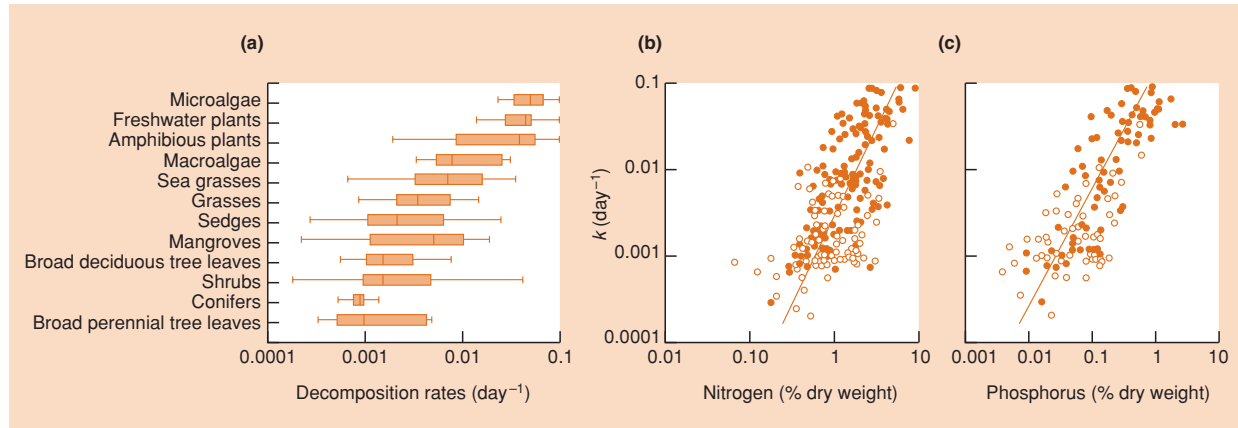


Figure 11.11 (a) Box plots showing the recorded decomposition rates of detritus from different sources. The decomposition rate is expressed as k (in log units per day), derived from the equation $W_t = W_0 e^{-kt}$, which describes the loss in plant dry weight (W) with time (t) since the initiation of measurements. Boxes encompass the 25 and 75% quartiles of all data from the literature for each plant type. The central line represents the median and bars extend to the 95% confidence limits. The relationships between decomposition rate and the initial concentrations in the tissues (% dry weight) of (b) nitrogen and (c) phosphorus are also shown. Solid lines represent fitted regression lines and open and closed circles represent detritus decomposing on land and submersed, respectively. (After Enriquez *et al.*, 1993.)

for phosphorus in freshwater communities, and either nitrogen or phosphorus in marine communities).

... competition and mutualism

Daufresne and Loreau (2001) developed a model that incorporates both mutualistic and competitive relationships and posed the question ‘what conditions must be met for plants and decomposers to coexist and for the ecosystem as a whole to persist?’ Their model showed that the plant–decomposer system is generally persistent (both plant and decomposer compartments reach a stable positive steady state) only if decomposer growth is limited by the availability of carbon in the detritus – and this condition can only be achieved if the competitive ability of the decomposers for a limiting nutrient (e.g. nitrogen) was great enough, compared to that of plants, to maintain themselves in a state of carbon limitation. When decomposers were not competitive enough, they became nutrient-limited and the system eventually collapsed. Daufresne and Loreau (2001) note that the few experimental studies so far performed show bacteria can, in fact, outcompete plants for inorganic nutrients.

In contrast to terrestrial plants, the bodies of animals have nutrient ratios that are of the same order as those of microbial biomass; thus their decomposition is not limited by the availability of nutrients, and animal bodies tend to decompose much faster than plant material.

When dead organisms or their parts decompose in or on soil, they begin to acquire the C : N ratio of the decomposers. On the whole, if material with a nitrogen content of less than 1.2–1.3% is added to soil, any available ammonium ions are absorbed. If the material has a nitrogen content greater than 1.8%, ammonium

ions tend to be released. One consequence is that the C : N ratios of soils tend to be rather constant around values of 10; the decomposer system is in general remarkably homeostatic. However, in extreme situations, where the soil is very acid or waterlogged, the ratio may rise to 17 (an indication that decomposition is slow).

It should not be thought that the only activity of the microbial decomposers of dead material is to respire away the carbon and mineralize the remainder. A major consequence of microbial growth is the accumulation of microbial by-products, particularly fungal cellulose and microbial polysaccharides, which may themselves be slow to decompose and contribute to maintaining soil structure.

11.3 Detritivore–resource interactions

11.3.1 Consumption of plant detritus

Two of the major organic components of dead leaves and wood are cellulose and lignin. These pose considerable digestive problems for animal consumers, most of which are not capable of manufacturing the enzymatic machinery to deal with them. Cellulose catabolism (cellulolysis) requires *cellulase* enzymes. Without these, detritivores are unable to digest the cellulose component of detritus, and so cannot derive from it either energy to do work or the simpler chemical modules to use in their own tissue synthesis. Cellulases of animal origin have been definitely identified in remarkably few species, including a cockroach and some higher termites in the subfamily Nasutitermitinae (Martin, 1991) and the shipworm *Teledo navalis*, a marine bivalve mollusc

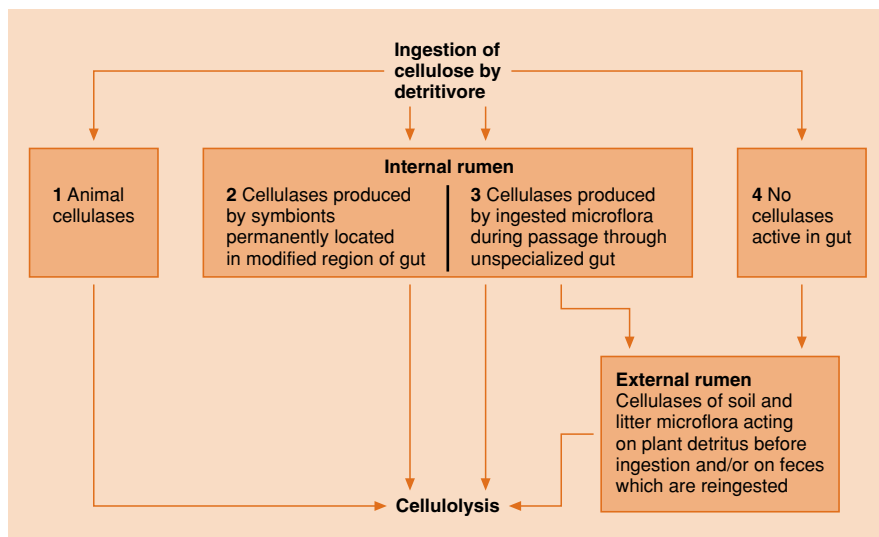


Figure 11.12 The range of mechanisms that detritivores adopt for digesting cellulose (cellulolysis). (After Swift *et al.*, 1979.)

that bores into the hulls of ships. In these organisms, cellulolysis poses no special problems.

most detritivores rely on microbial cellulases – they do not have their own

The majority of detritivores, lacking their own cellulases, rely on the production of cellulases by associated decomposers or, in some cases, protozoa. The interactions range from *obligate mutualism* between a detritivore and a specific and permanent gut microflora or microfauna, through *facultative mutualism*, where the animals make use of cellulases produced by a microflora that is ingested with detritus as it passes through an unspecialized gut, to animals that ingest the metabolic products of external cellulase-producing microflora associated with decomposing plant remains or feces (Figure 11.12).

woodlice rely on ingested microbial organisms

A wide range of detritivores appear to have to rely on the exogenous microbial organisms to digest cellulose. The invertebrates then consume the partially digested plant detritus along with its associated bacteria and fungi, no doubt obtaining a significant proportion of the necessary energy and nutrients by digesting the microflora itself. These animals, such as the spring-tail *Tomocerus*, can be said to be making use of an ‘external rumen’ in the provision of assimilable materials from indigestible plant remains. This process reaches a pinnacle of specialization in ambrosia beetles and in certain species of ants and termites that ‘farm’ fungus in specially excavated gardens (see Chapter 13).

cockroaches and termites rely on bacteria and protozoa

Clear examples of obligate mutualism are found amongst certain species of cockroach and termite that rely on symbiotic bacteria or protozoa for the digestion of structural plant polysaccharides. Nalepa *et al.* (2001) describe

the evolution of digestive mutualisms among the Dictyoptera (cockroaches and termites) from cockroach-like ancestors in the Upper Carboniferous that fed on rotting vegetation and relied on an ‘external rumen’. The next stages involved progressive internalization of the microbiota associated with plant detritus, from indiscriminate coprophagy (feeding on feces of a variety of detritivorous species) through increasing levels of gregarious and social behavior that ensured neonates received appropriate inocula of gut biota. When proctodeal trophallaxis (the direct transfer of hindgut fluids from the rectal pouch of the parent to the mouth of the newborn young) evolved in certain cockroaches and lower termites, some microbes were captured and became ecologically dependent on the host. This specialized state ensured the direct transfer of the internal rumen, particularly those components that would degenerate if exposed to the external environment. In lower termites, such as *Eutermes*, symbiotic protozoa may make up more than 60% of the insect’s body weight. The protozoa are located in the hindgut, which is dilated to form a rectal pouch. They ingest fine particles of wood, and are responsible for extensive cellulolytic activity, though bacteria are also implicated. Termites feeding on wood generally show effective digestion of cellulose but not of lignin, except for *Reticulitermes*, which has been reported to digest 80% or more of the lignin present in its food.

Given the versatility apparent in the evolutionary process, it may seem surprising that so few animals that consume plants can produce their own cellulase enzymes. Janzen (1981) has argued that cellulose is the master construction material of plants ‘for the same reason that we construct houses of concrete in areas of high termite activity’. He views the use of cellulose, therefore, as a defense against attack, since higher organisms can rarely digest it unaided. From a different perspective,

why no animal cellulases?

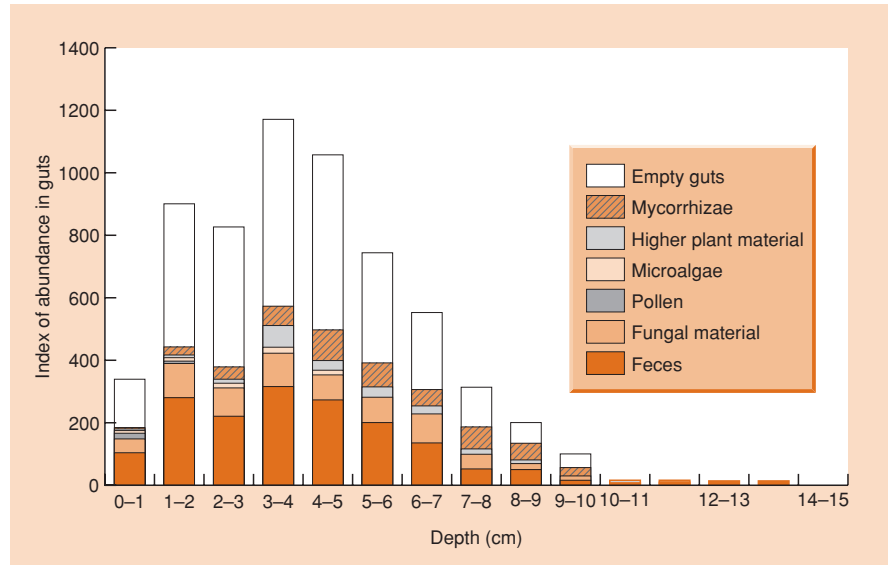


Figure 11.13 The distribution of gut content categories of springtails ($n = 6255$) (Collembola; all species combined) in relation to depth in the litter/soil of beech forests in Belgium. (After Ponge, 2000.)

it has been suggested that cellulolytic capacity is uncommon simply because it is a trait that is rarely advantageous for animals to possess (Martin, 1991). For one thing, diverse bacterial communities are commonly found in hindguts and this may have facilitated the evolution of symbiont-mediated cellulolysis. For another, the diets of plant-eaters generally suffer from a limited supply of critical nutrients, such as nitrogen and phosphorus, rather than of energy, which cellulolysis would release. This imposes the need for processing large volumes of material to extract the required quantities of nutrients, rather than extracting energy efficiently from small volumes of material.

Because microbes, plant detritus and animal feces are often very intimately associated, there are inevitably many generalist consumers that ingest all these resources. In other words, many animals simply cannot manage to take a mouthful of one without the others. Figure 11.13 shows the various components of the gut contents of 45 springtail species (all species combined) collected at different depths in the litter and soil of beech forests in Belgium. Species that occurred in the top 2 cm lived in a habitat derived from beech leaves at various stages of microbial decomposition where microalgae, feces of slugs and woodlice, and pollen grains were also common. Their diets contained all the local components but little of the very abundant beech litter. At intermediate depths (2–4 cm) the springtails ate mainly spores and hyphae of fungi together with invertebrate feces (particularly the freshly deposited feces of enchytraeid pot worms). At the lowest depths, their diets consisted mainly of mycorrhizal material (the springtails browsed the fungal part of the fungal/plant root assemblage) and higher plant detritus (mainly derived from plant roots). There were clear interspecific differences

detritus and microbial organisms are typically consumed together

in both depth distributions and the relative importance of the different dietary components, and some species were more specialized feeders than others (e.g. *Isotomiella minor* ate only feces whereas *Willemia aspinata* ate only fungal hyphae). But most consumed more than one of the potential diet components and many were remarkably generalist (e.g. *Protaphorura eichhorni* and *Mesaphorura yosii*) (Ponge, 2000).

11.3.2 Consumption of fallen fruit

Of course, not all plant detritus is so difficult for detritivores to digest. Fallen fruit, for example, is readily exploited by many kinds of opportunist feeders, including insects, birds and mammals. However, like all detritus, decaying fruits have associated with them a microflora, in this case mainly dominated by yeasts. Fruit-flies (*Drosophila* spp.) specialize at feeding on these yeasts and their by-products; and in fruit-laden domestic compost heaps in Australia, five species of fruit-fly show differing preferences for particular categories of rotting fruit and vegetables (Oakeshott *et al.*, 1982). *Drosophila hydei* and *D. immigrans* prefer melons, *D. busckii* specializes on rotting vegetables, while *D. simulans* is catholic in its tastes for a variety of fruits. The common *D. melanogaster*, however, shows a clear preference for rotting grapes and pears. Note that rotting fruits can be highly alcoholic. Yeasts are commonly the early colonists and the fruit sugars are fermented to alcohol, which is normally toxic, eventually even to the yeasts themselves. *D. melanogaster* tolerates such high levels of alcohol because it produces large quantities of alcohol dehydrogenase (ADH), an enzyme that breaks down ethanol to harmless metabolites. Decaying vegetables produce

fruit-flies and rotten fruit

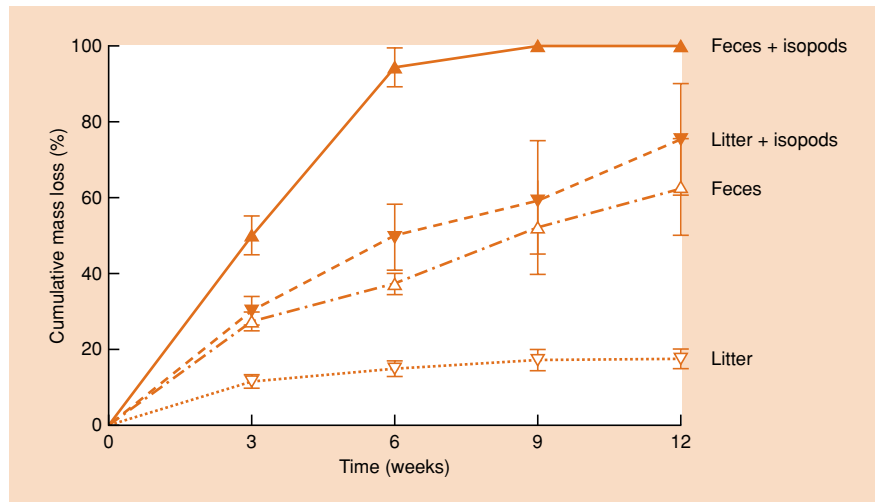


Figure 11.14 The cumulative mass loss of beech leaf litter and feces of grazing caterpillars (*Operophtera fagata*) in the presence and absence of feeding by isopods. Standard errors are shown. (After Zimmer & Topp, 2002.)

little alcohol, and *D. busckii*, which is associated with them, produces very little ADH. Intermediate levels of ADH were produced by the species preferring moderately alcoholic melons. The boozy *D. melanogaster* is also associated with winery wastes!

11.3.3 Feeding on invertebrate feces

isopods do best when they can eat their own feces

A large proportion of dead organic matter in soils and aquatic sediments may consist of invertebrate feces, which generalist detritivores often include in their diets. Some of the feces derive from grazing insects. In the laboratory, the feces of caterpillars of *Operophtera fagata* that had grazed leaves of beech (*Fagus sylvatica*) under the influence of leaching and microbial degradation decomposed faster than leaf litter itself; however, the decomposition rate was much enhanced when detritivorous isopods (*Porcellio scaber* and *Oniscus asellus*) fed on the feces (Figure 11.14). Thus, rates of decomposition and nutrient release into the soil from grazer feces can be increased through the feeding activity of coprophagous detritivores.

'coprophagy' may be more valuable when detrital quality is low

Feces of detritivores are common in many environments. In some cases, reingestion of feces may be critically important, by providing essential micro-nutrients or highly assimilable resources. In most cases, however, there are probably not marked nutritive benefits of feeding on feces compared with the detritus from which the feces were derived. Thus, the isopod *Porcellio scaber* gained no more from feeding on its feces, even when these were experimentally inoculated with microbes, than from feeding directly on the leaf litter of alder (*Alnus glutinosa*) (Kautz *et al.*, 2002). On the other hand, in the case of the less nutritionally preferred leaves of oak (*Quercus robur*), inoculated feces provided a small but

significant increase in growth rate compared to the parent oak leaf material. Coprophagy may be more valuable when detrital quality is particularly low.

a midge and a cladoceran eat each other's feces

A remarkable story of coprophagy was unraveled in some small bog lakes in northeast England (MacLachlan *et al.*, 1979). These murky water bodies have restricted light penetration because of dissolved humic substances derived from the surrounding sphagnum peat, and they are characteristically poor in plant nutrients. Primary production is insignificant. The main organic input consists of poor-quality peat particles resulting from wave erosion of the banks. By the time the peat has settled from suspension it has been colonized, mainly by bacteria, and its caloric and protein contents have increased by 23 and 200%, respectively. These small particles are consumed by *Chironomus lugubris* larvae, the detritivorous young of a nonbiting chironomid midge. The feces the larvae produce become quite richly colonized by fungi, microbial activity is enhanced, and they would seem to constitute a high-quality food resource. But they are not reingested by *Chironomus* larvae, mainly because they are too large and too tough for its mouthparts to deal with. However, another common inhabitant of the lake, the small crustacean *Chydorus sphaericus*, finds chironomid feces very attractive. It seems always to be associated with them and probably depends on them for food. *Chydorus* clasps the chironomid fecal pellet just inside the valve of its carapace and rotates it while grazing the surface, causing gradual disintegration. In the laboratory, the presence of chydorids has been shown to speed up dramatically the breakdown of large *Chironomus* pellets to smaller particles. The final and most intriguing twist to the story is that the fragmented chironomid feces (mixed probably with chydorid feces) are now small enough to be used again by *Chironomus*. It is probable that *Chironomus lugubris* larvae grow faster when in the presence of *Chydorus sphaericus* because of

the availability of suitable fecal material to eat. The interaction benefits both participants.

11.3.4 Feeding on vertebrate feces

carnivore dung is attacked mainly by bacteria and fungi

The dung of carnivorous vertebrates is relatively poor-quality stuff. Carnivores assimilate their food with high efficiency (usually 80% or more is digested) and

their feces retain only the least digestible components. In addition, carnivores are necessarily much less common than herbivores, and their dung is probably not sufficiently abundant to support a specialist detritivore fauna. What little research has been done suggests that decay is effected almost entirely by bacteria and fungi (Putman, 1983).

'autocoprophagy' among mammalian herbivores

In contrast, herbivore feces still contain an abundance of organic matter. Autocoprophagy (reingesting one's own feces) is quite a widespread habit among small to medium-sized mam-

malian herbivores, being reported from rabbits and hares, rodents, marsupials and a primate (Hirakawa, 2001). Many species produce soft and hard feces, and it is the soft feces that are usually reingested (directly from the anus), being rich in vitamins and microbial protein. If prevented from reingestion, many animals exhibit symptoms of malnutrition and grow more slowly.

herbivore dung supports its own characteristic detritivores

Herbivore dung is also sufficiently thickly spread in the environment to support its own characteristic fauna, consisting of many occasional visitors but with several specific dung-feeders.

Dung removal varies both seasonally and spatially. In tropical and in warm temperate regions most activity occurs during summer rainfall, whereas in Mediterranean-type climates dung removal is highest during spring after the winter rainfall and again in mid-summer when temperatures are high (Davis, 1996). Dung removal also occurs at greater rates in unshaded situations and is faster on sand than on harder, more compacted clay soils (Davis, 1996). A wide range of animals are involved, including earthworms, termites and, in particular, beetles.

A good example of the predominant role of beetles is provided by elephant dung. Two main patterns of decay can be recognized, related to the wet and dry seasons. During the rains, within a few minutes of dung deposition the area is alive with beetles. The adult dung beetles feed on the dung but they also bury large quantities along with their eggs to provide food for the developing larvae. For example, the large African dung beetle, *Heliocopris dilloni*, carves a lump out of fresh dung and rolls this away for burying several meters from the original dung pile. Each beetle buries sufficient dung for several eggs. Once underground, a

small quantity of dung is shaped into a cup, and lined with soil; a single egg is laid and then more dung is added to produce a sphere that is almost entirely covered with a thin layer of soil. A small area at the top of the ball, close to the location of the egg, is left clear of soil, possibly to facilitate gas exchange. After hatching, the larva feeds by a rotating action in the dung ball, excavating a hollow, and, incidentally, feeding on its own feces as well as the elephant's (Figure 11.15). When all the food supplied by its parents is used up, the larva covers the inside of its cell with a paste of its own feces, and pupates.

The full range of tropical dung beetles in the family Scarabaeidae vary in size from a few millimeters in length up to the 6 cm long *Heliocopris*. Not all

a diversity of dung beetles

remove dung and bury it at a distance from the dung pile. Some excavate their nests at various depths immediately below the pile, while others build nest chambers within the dung pile itself. Beetles in other families do not construct chambers but simply lay their eggs in the dung, and their larvae feed and grow within the dung mass until fully developed, when they move away to pupate in the soil. The beetles associated with elephant dung in the wet season may remove 100% of the dung pile. Any left may be processed by other detritivores such as flies and termites, as well as by decomposers.

Dung that is deposited in the dry season is colonized by relatively few beetles (adults emerge only in the rains). Some microbial activity is evident but this soon declines as the feces dry out. Rewetting during the rains stimulates more microbial activity but beetles do not exploit old dung. In fact a dung pile deposited in the dry season may persist for longer than 2 years, compared with 24 h or less for one deposited during the rains.

Bovine dung has provided an extraordinary and economically very important problem in Australia. During the past two centuries the cow population

Australian cow dung poses a problem

increased from just seven individuals (brought over by the first English colonists in 1788) to 30 million or so. These produce some 300 million dung pats per day, covering as much as 6 million acres per year with dung. Deposition of bovine dung poses no particular problem elsewhere in the world, where bovines have existed for millions of years and have an associated fauna that exploits the fecal resources. However, the largest herbivorous animals in Australia, until European colonization, were marsupials such as kangaroos. The native detritivores that deal with the dry, fibrous dung pellets that these leave cannot cope with cow dung, and the loss of pasture under dung has imposed a huge economic burden on Australian agriculture. The decision was therefore made in 1963 to establish in Australia beetles of African origin, able to dispose of bovine dung in the most important places and under the most prevalent conditions where cattle are raised (Waterhouse, 1974); more than 20 species have been introduced (Doube *et al.*, 1991).

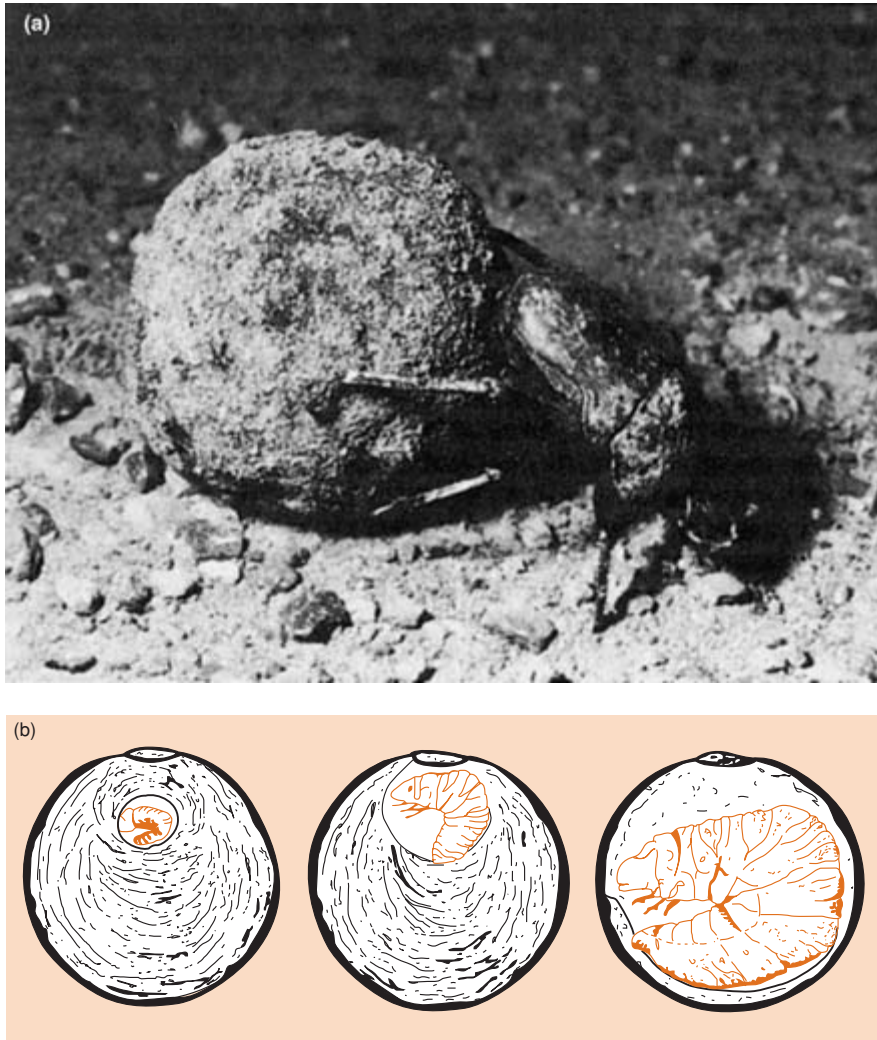


Figure 11.15 (a) An African dung beetle rolling a ball of dung. (Courtesy of Heather Angel.) (b) The larva of the dung beetle *Heliocopris* excavates a hollow as it feeds within the dung ball. (After Kingston & Coe, 1977.)

Adding to the problem, Australia is plagued by native bushflies (*Musca vetustissima*) and buffalo flies (*Haematobia irritans exigua*) that deposit eggs on dung pats. The larvae fail to survive in dung that has been buried by beetles, and the presence of beetles has been shown to be effective at reducing fly abundance (Tyndale-Biscoe & Vogt, 1996). Success depends on dung being buried within about 6 days of production, the time it takes for the fly egg (laid on fresh dung) to hatch and develop to the pupal stage. Edwards and Aschenborn (1987) surveyed the nesting behavior in southern Africa of 12 species of dung beetles in the genus *Onitis*. They concluded that *O. uncinatus* was a prime candidate for introduction to Australia for fly-control purposes, since substantial amounts of dung were buried on the first night after pad colonization. The least suitable species, *O. viridulus*, spent several days constructing a tunnel and did not commence burying until 6–9 days had elapsed.

11.3.5 Consumption of carrion

When considering the decomposition of dead bodies, it is helpful to distinguish three categories of organisms that attack carcasses. As before, both decomposers

many carnivores are opportunistic carrion-feeders . . .

and invertebrate detritivores have a role to play. For example, the tenebrionid beetles *Argoporis apicalis* and *Cryptadius tarsalis* are particularly abundant on islands in the Gulf of California where large colonies of seabirds nest; here they feed on bird carcasses, as well as fish debris associated with the bird colonies (Sanchez-Pinero & Polis, 2000). In the case of carrion feeding, however, scavenging vertebrates are often also of considerable importance. Many carcasses of a size to make a single meal for one of a few of these scavenging detritivores will be removed completely within a very short time of death, leaving nothing for bacteria,

fungi or invertebrates. This role is played, for example, by arctic foxes and skuas in polar regions, by crows, gluttons and badgers in temperate areas, and by a wide variety of birds and mammals, including kites, jackals and hyenas, in the tropics.

... and vice versa

The chemical composition of the diet of carrion-feeders is quite distinct from that of other detritivores, and this is reflected in their complement of enzymes. Carbohydrase activity is weak or absent, but protease and lipase activity is vigorous. Carrion-feeding detritivores possess basically the same enzymatic machinery as carnivores, reflecting the chemical identity of their food. In fact, many species of carnivore (such as lions, *Panthera leo*) are also opportunistic carrion-feeders (DeVault & Rhodes, 2002) whilst classic carrion-feeders such as hyenas (*Crocuta crocuta*) sometimes operate as carnivores.

the arctic fox: a facultative carrion-feeder

Arctic foxes (*Alopex lagopus*) illustrate how the diet of facultative carrion-feeders can vary with food availability. Lemmings (*Dicrostonyx* and *Lemmus* spp.) are the live prey of foxes over much of their range and for much of the time (Elmhagen *et al.*, 2000). However, lemming populations go through dramatic population cycles (see Chapter 14), forcing the foxes to switch to alternative foods such as migratory birds and their eggs (Samelius & Alisauskas, 2000). In winter, marine foods become available when foxes can move onto the sea ice and scavenge carcasses of seals killed by polar bears. Roth (2002) investigated the extent to which foxes switched to carrion feeding in winter by comparing the ratios of carbon isotopes ($^{13}\text{C} : ^{12}\text{C}$) of suspected food (marine organisms have characteristically higher ratios than terrestrial organisms) and of fox hair (since carbon isotope signatures of predator tissue reflect the ratios of the prey consumed). Figure 11.16 shows that in three of the 4 years of the study the isotope signature of fox hair samples was much increased in winter, as expected if seal carrion was a major component of the diet. In the winter of 1994, however, a marked shift was not evident and it is of interest that lemming density was high at this time. It seems that foxes switched to seal carrion when the formation of sea ice made this possible, but only when alternative prey were not available.

seasonal variation in invertebrate and microbial activity

The relative roles played by decomposers, invertebrates and vertebrates are influenced by factors that affect the speed with which carcasses are discovered by scavengers in relation to the rate at which they disappear through microbial and invertebrate activity. This is illustrated for small rodent carcasses whose disappearance/decomposition was monitored in the Oxfordshire countryside in both the summer–fall and winter–spring periods (Figure 11.17). There are two points to note. First, the rate at which carcasses were removed was faster during the summer and fall, reflecting a greater scavenger activity at this time (presumably

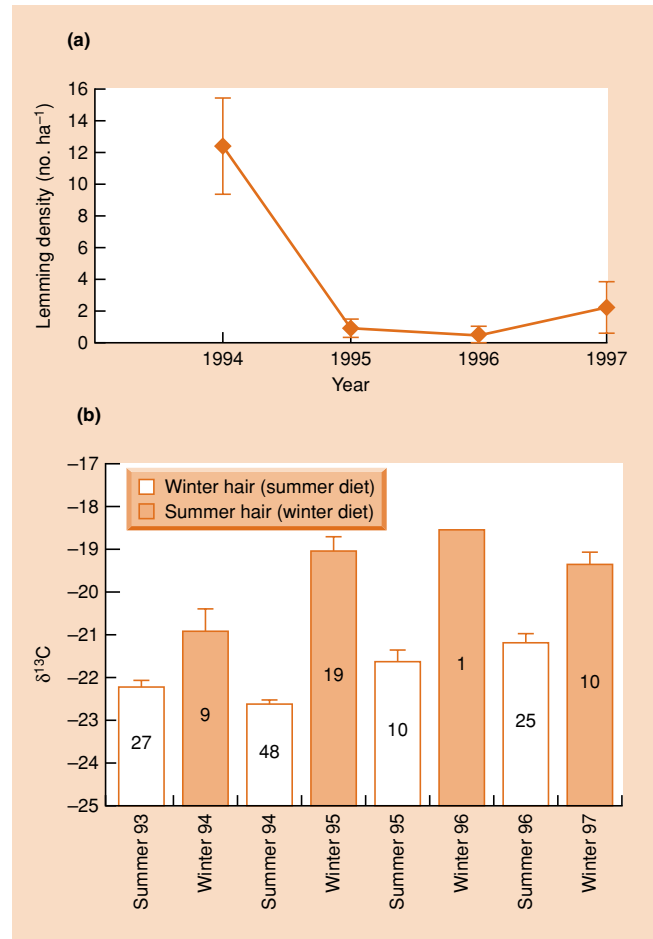


Figure 11.16 (a) Annual changes in lemming density in the summer, near Cape Churchill in Manitoba, Canada, and (b) carbon isotope ratios (mean ± SE) of fox hair in the winter (reflecting summer diet) and in the summer (reflecting winter diet). Numbers on the bars indicate sample sizes. (After Roth, 2002.)

because of higher scavenger population densities and/or higher feeding rates – these were not monitored in the study). Secondly, a greater percentage of the rodent bodies were removed in the winter–spring period, albeit over a longer timescale. At a time when microbial decay proceeds most slowly, all the carcasses persisted for long enough to be found by scavengers. During the summer and fall, decomposition was much more rapid and any carcass that was undiscovered for 7 or 8 days would have been largely decomposed and removed by bacteria, fungi and invertebrate detritivores.

Certain components of animal corpses are particularly resistant to attack and are the slowest to disappear. However, some consumer species possess the enzymes to deal with them. For example, the blowfly larvae of *Lucilia* species produce a collagenase that can digest the collagen

specialist consumers of bone, hair and feathers

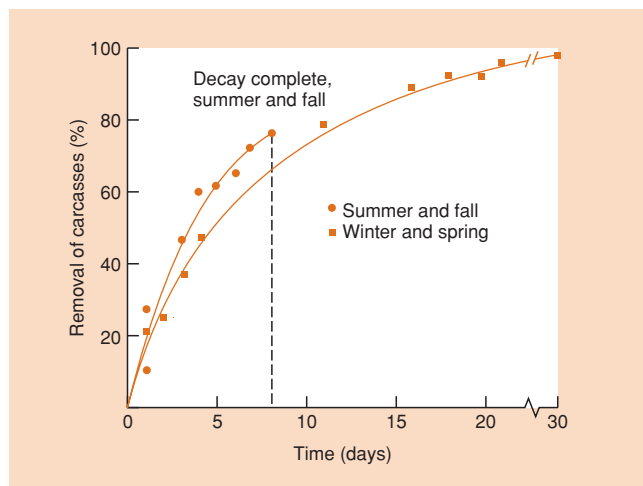


Figure 11.17 The rate of removal of small mammal carcasses in the Oxfordshire (UK) countryside in two periods: summer–fall and winter–spring. (After Putman, 1983.)

and elastin present in tendons and soft bones. The chief constituent of hair and feathers, keratin, forms the basis of the diet of species characteristic of the later stages of carrion decomposition, in particular tineid moths and dermestid beetles. The midgut of these insects secretes strong reducing agents that break the

resistant covalent links binding together peptide chains in the keratin. Hydrolytic enzymes then deal with the residues. Fungi in the family Onygenaceae are specialist consumers of horn and feathers. It is the corpses of larger animals that generally provide the widest variety of resources and thus attract the greatest diversity of carrion consumers (Doube, 1987). In contrast, the carrion community associated with dead snails and slugs consists of a relatively small number of sarcophagid and calliphorid flies (Kneidel, 1984).

One group of carrion-feeding invertebrates deserves special attention – the burying beetles (*Nicrophorus* spp.) (Scott, 1998). These species live exclusively on carrion on which they play out their extraordinary life history. Adult *Nicrophorus*, using their sensitive chemoreceptors, arrive at the carcass of a small mammal or bird within an hour or two of death. The beetle may tear flesh from the corpse and eat it or, if decomposition is sufficiently advanced, consume blowfly larvae instead. However, should a burying beetle arrive at a completely fresh corpse it sets about burying it where it lies, or may drag the body (many times its own weight) for several meters before starting to dig. It works beneath the corpse, painstakingly excavating and dragging the small mammal down little by little until it is completely underground (Figure 11.18). The various species of *Nicrophorus* vary in body size (and thus the size of corpse utilized), reproductive period (and

remarkable burying
beetles

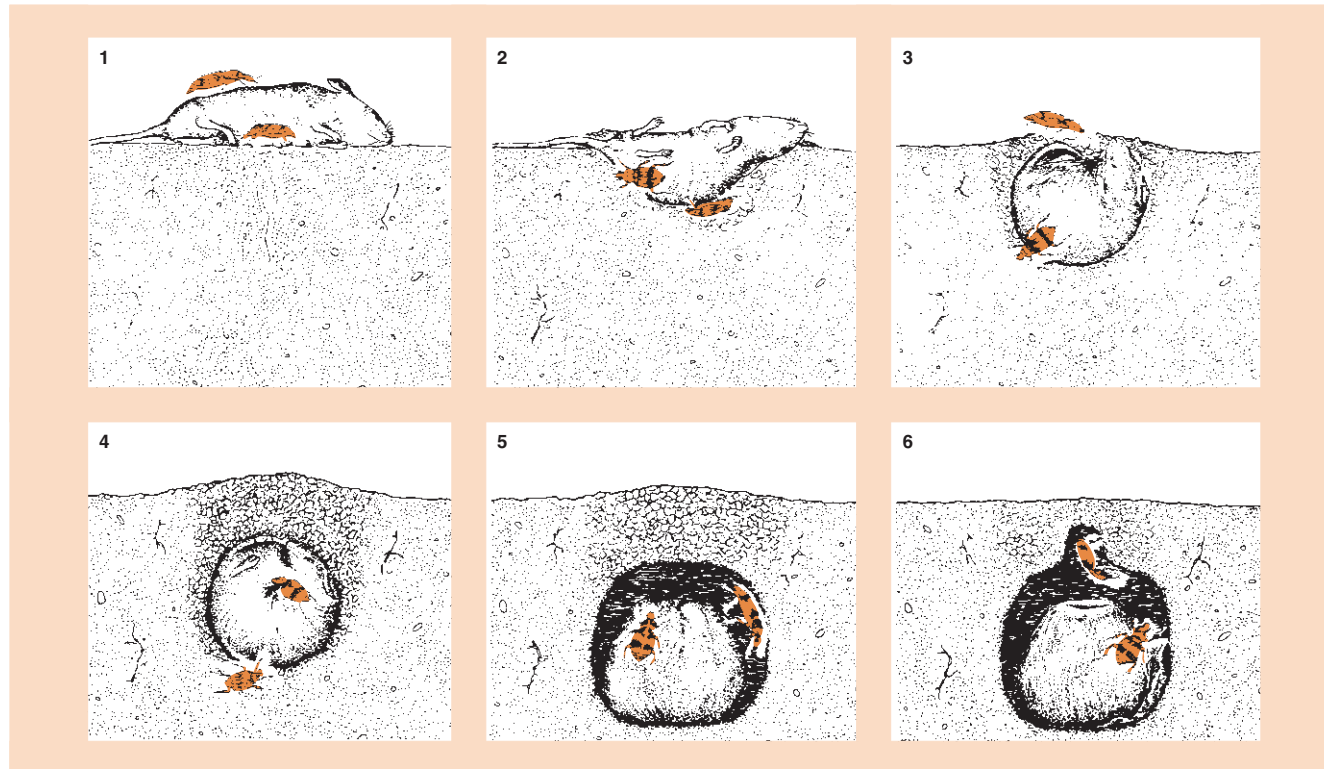


Figure 11.18 Burial of a mouse by a pair of *Nicrophorus* beetles. (After Milne & Milne, 1976.)

thus the season of activity), diel activity (some are diurnal, some crepuscular and some nocturnal) and the habitat they use (coniferous forest, hardwood forest, field, marsh or generalist) (Scott, 1998). Some species, such as *N. vespilloides*, only just cover the corpse, while others, including *N. germanicus*, may bury it to a depth of 20 cm. During the excavation, other burying beetles are likely to arrive. Competing individuals of the same or other species are fiercely repulsed, sometimes leading to the death of one combatant. A prospective mate, on the other hand, is accepted and the male and female work on together.

The buried corpse is much less susceptible to attack by other invertebrates than it was while on the surface. Additional protection is provided, under some circumstances, by virtue of a mutualistic relationship between the beetles and a species of mite, *Poecilochirus necrophori*, which invariably infests adult burying beetles, hitching a ride to a suitable carrion source. When the carcass is first buried the beetle systematically removes its hair and this clears it of virtually all the eggs of blowflies. However, if the carcass is buried only shallowly, flies will often lay more eggs and maggots will compete with the beetle larvae. It is now that the presence of mites has a beneficial effect. By piercing and consuming the fly eggs, the mites keep the carcass free of the beetle's competitors and dramatically improve beetle brood success (Wilson, 1986). Both adults, or sometimes just the female, remain in the chamber and provide parental care. A conical depression is prepared in the top of the meat-ball, into which droplets of partially digested meat are regurgitated. Older larvae are able to feed themselves but only when their offspring are ready to pupate do the adults force their way out through the soil and fly away.

carrion feeders on the sea bed

We have already noted that in freshwater environments carrion lack a specialized fauna. However, specialist carrion-feeders are found on the sea bed in very deep parts of the oceans. As the detritus sinks through very deep water all but the largest particles of organic matter are completely decomposed before they reach the bottom. In contrast, the occasional body of a fish, mammal or large invertebrate does settle on the sea bed. A remarkable diversity of scavengers exist there, though at low density, and these possess several characteristics that match a way of life in which meals are well spread out in space and time. For example, Dahl (1979) described several genera of deep sea gammarid crustaceans which, unlike their relatives at shallower depths and in fresh water, possess dense bundles of exposed chemosensory hairs that sense food, and sharp mandibles that can take large bites from carrion. These animals also have the capacity to gorge themselves far beyond what is normal in amphipods. Thus *Paralicella* possesses a soft body wall that can be stretched when feeding on a large meal so that the animal swells to two or three times its normal size, and *Hirondella* has a midgut that expands to fill almost the entire abdominal cavity and in which it can store meat.

11.4 Conclusion

Decomposer communities are, in their composition and activities, as diverse as or more diverse than any of the communities more commonly studied by ecologists. Generalizing about them is unusually difficult because the range of conditions experienced in their lives is so varied. As in all natural communities, the inhabitants not only have specialized requirements for resources and conditions, but their activities change the resources and conditions available for others. Most of this happens hidden from the view of the observer, in the crevices and recesses of soil and litter and in the depths of water bodies.

Despite these difficulties, some broad generalizations may be made.

- 1 Decomposers and detritivores tend to have low levels of activity when temperatures are low, aeration is poor, soil water is scarce and conditions are acid.
- 2 The structure and porosity of the environment (soil or litter) is of crucial importance, not only because it affects the factors listed in point 1 but because many of the organisms responsible for decomposition must swim, creep, grow or force their way through the medium in which their resources are dispersed.
- 3 The activities of the decomposers and detritivores are intimately interlocked, and may in some cases be synergistic. For this reason, it is very difficult to unravel their relative importance in the decomposition process.
- 4 Many of the decomposers and detritivores are specialists and the decay of dead organic matter results from the combined activities of organisms with widely different structures, forms and feeding habits.
- 5 Organic matter may cycle repeatedly through a succession of microhabitats within and outside the guts and feces of different organisms, as they are degraded from highly organized structures to their eventual fate as carbon dioxide and mineral nutrients.
- 6 The activity of decomposers unlocks the mineral resources such as phosphorus and nitrogen that are fixed in dead organic matter. The speed of decomposition will determine the rate at which such resources are released to growing plants (or become free to diffuse and thus to be lost from the ecosystem). This topic is taken up and discussed in Chapter 18.
- 7 Many dead resources are patchily distributed in space and time. An element of chance operates in the process of their colonization; the first to arrive have a rich resource to exploit, but the successful species may vary from dung pat to dung pat, and from corpse to corpse. The dynamics of competition between exploiters of such patchy resources require their own particular mathematical models (see Chapter 8). Because detritus is often an 'island' in a sea of quite different habitat, its study is conceptually similar to that discussed in Chapter 21 under the heading of island biogeography (see Section 21.5).

8 Finally, it may be instructive at this point to switch the emphasis away from the success with which decomposers and detritivores deal with their resources. It is, after all, the failure of organisms to decompose wood rapidly that makes the existence of forests possible! Deposits of peat, coal and oil are further testaments to the failures of decomposition.

Summary

We distinguish two groups of organisms that make use of dead organic matter (saprotrophs): decomposers (bacteria and fungi) and detritivores (animal consumers of dead matter). These do not control the rate at which their resources are made available or regenerate; they are dependent on the rate at which some other force (senescence, illness, fighting, the shedding of leaves by trees) releases the resource on which they live. They are donor controlled. Nevertheless, it is possible to see an indirect 'mutualistic' effect through the release of nutrients from decomposing litter, which may ultimately affect the rate at which trees produce more litter.

Immobilization occurs when an inorganic nutrient element is incorporated into an organic form – primarily during the growth of green plants. Conversely, decomposition involves the release of energy and the *mineralization* of chemical nutrients – the conversion of elements from an organic to inorganic form. Decomposition is defined as the gradual disintegration of dead organic matter and is brought about by both physical and biological agencies. It culminates, often after a reasonably predictable succession of colonizing decomposers, with complex

energy-rich molecules being broken down into carbon dioxide, water and inorganic nutrients.

Most microbial decomposers are quite specialized, as are the tiny consumers of bacteria and fungi (microbivores), but detritivores are more often generalists. The larger the detritivore, the less able it is to distinguish between microbes as food and the detritus on which these are growing. We discuss the relative roles in decomposition of decomposers and detritivores in terrestrial, freshwater and marine environments.

The rate at which dead organic matter decomposes is strongly dependent on its biochemical composition and on the availability of mineral nutrients in the environment. Two of the major organic components of dead leaves and wood are cellulose and lignin. These pose considerable digestive problems for animal consumers, most of which are not capable of manufacturing the enzymatic machinery to deal with them. Most detritivores depend on microbial organisms to digest cellulose, in a variety of increasingly intimate associations. Dead fruit is a lot easier for detritivores to deal with.

Feces and carrion are abundant dead organic resources in all environments and, once again, microbial organisms and detritivores both play important roles. Many detritivores feed on feces, and the dung of herbivores (but not carnivores) supports its own characteristic fauna. Similarly, many carnivores are opportunistic feeders on carrion but there is also a specialized carrion-feeding fauna.

Decomposer communities are, in their composition and activities, as diverse as or more diverse than any of the communities more commonly studied by ecologists.

Chapter 12

Parasitism and Disease



12.1 Introduction: parasites, pathogens, infection and disease

Previously, in Chapter 9, we defined a parasite as an organism that obtains its nutrients from one or a very few host individuals, normally causing harm but not causing death immediately. We must follow this now with some more definitions, since there are a number of related terms that are often misused, and it is important not to do so.

When parasites colonize a host, that host is said to harbor an *infection*. Only if that infection gives rise to symptoms that are clearly harmful to the host should the host be said to have a *disease*. With many parasites, there is a presumption that the host can be harmed, but no specific symptoms have as yet been identified, and hence there is no disease. ‘*Pathogen*’ is a term that may be applied to any parasite that gives rise to a disease (i.e. is ‘pathogenic’). Thus, measles and tuberculosis are infectious diseases (combinations of symptoms resulting from infections). Measles is the result of a measles virus infection; tuberculosis is the result of a bacterial (*Mycobacterium tuberculosis*) infection. The measles virus and *M. tuberculosis* are pathogens. But measles is not a pathogen, and there is no such thing as a tuberculosis infection.

Parasites are an important group of organisms in the most direct sense. Millions of people are killed each year by various types of infection, and many millions more are debilitated or deformed (250 million cases of elephantiasis at present, over 200 million cases of bilharzia, and the list goes on). When the effects of parasites on domesticated animals and crops are added to this, the cost in terms of human misery and economic loss becomes immense. Of course, humans make things easy for the parasites by living in dense and aggregated populations and forcing their domesticated animals and crops to do the same. One of the key questions we will address in this chapter is: ‘to what extent are animals and plant populations *in general* affected by parasitism and disease?’

Parasites are also important numerically. An organism in a natural environment that does *not* harbor several species of parasite

is a rarity. Moreover, many parasites and pathogens are host-specific or at least have a limited range of hosts. Thus, the conclusion seems unavoidable that more than 50% of the species on the earth, and many more than 50% of individuals, are parasites.

12.2 The diversity of parasites

The language and jargon used by plant pathologists and animal parasitologists are often very different, and there are important differences in the ways in which animals and plants serve as habitats for parasites, and in the way they respond to infection. But for the ecologist, the differences are less striking than the resemblances, and we therefore deal with the two together. One distinction that is useful, though, is that between microparasites and macroparasites (Figure 12.1) (May & Anderson, 1979).

Microparasites are small and often intracellular, and they multiply directly within their host where they are often extremely numerous. Hence, it is generally difficult, and usually inappropriate, to estimate precisely the number of microparasites in a host. The number of infected hosts, rather than the number of parasites, is the parameter usually studied. For example, a study of a measles epidemic will involve counting the number of cases of the disease, rather than the number of particles of the measles virus.

Macroparasites have a quite different biology: they grow but do not multiply in their host, and then produce specialized infective stages (microparasites do not do this) that are released to infect new hosts. The macroparasites of animals mostly live on the body or in the body cavities (e.g. the gut), rather than within the host cells. In plants, they are generally intercellular. It is usually possible to count or at least estimate the numbers of macroparasites in or on a host (e.g. worms in an intestine or lesions on a leaf), and the numbers of parasites as well as the numbers of infected hosts can be studied by the epidemiologist.

micro- and
macroparasites

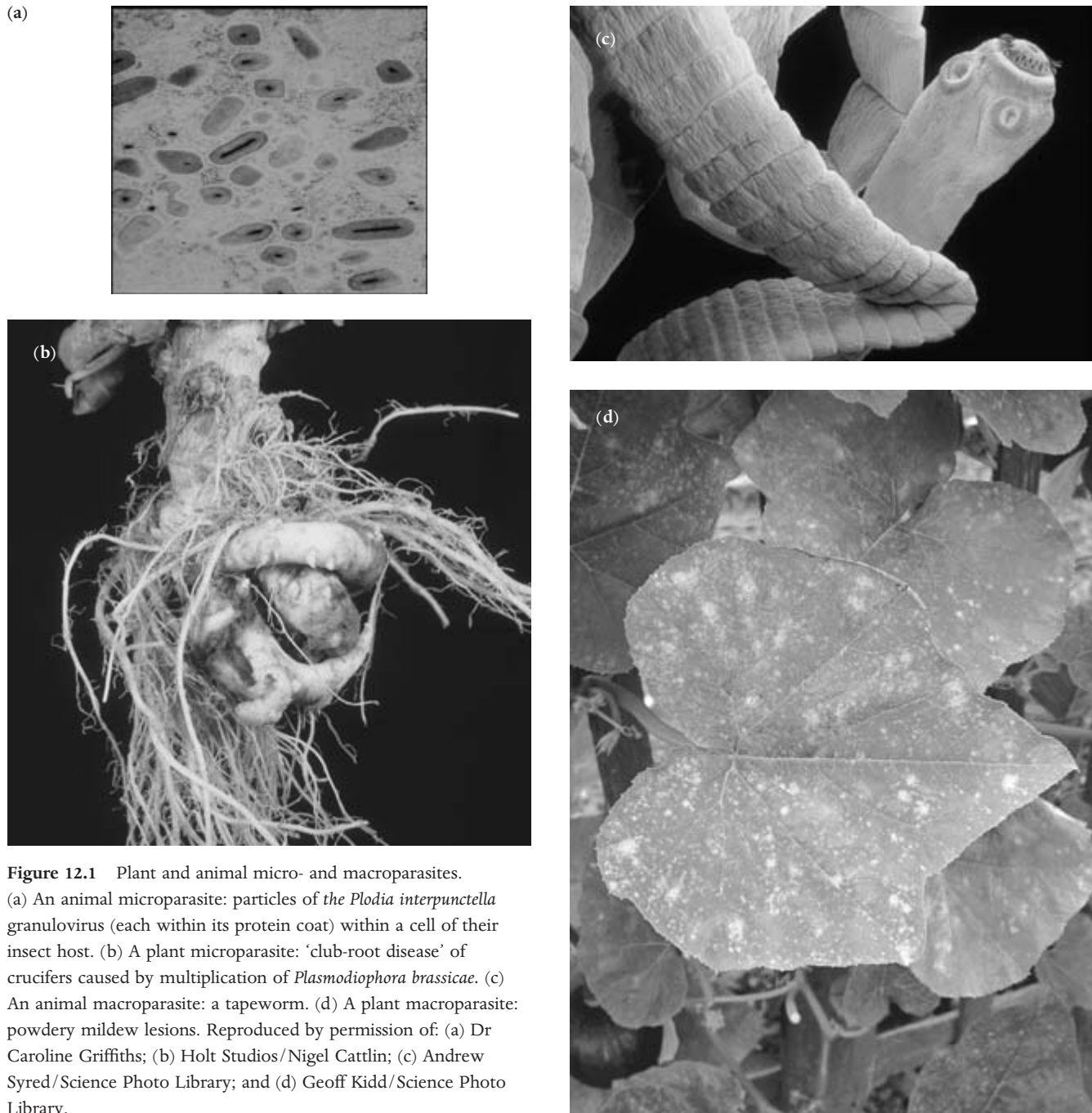


Figure 12.1 Plant and animal micro- and macroparasites. (a) An animal microparasite: particles of the *Plodia interpunctella* granulovirus (each within its protein coat) within a cell of their insect host. (b) A plant microparasite: ‘club-root disease’ of crucifers caused by multiplication of *Plasmodiophora brassicae*. (c) An animal macroparasite: a tapeworm. (d) A plant macroparasite: powdery mildew lesions. Reproduced by permission of: (a) Dr Caroline Griffiths; (b) Holt Studios/Nigel Cattlin; (c) Andrew Syred/Science Photo Library; and (d) Geoff Kidd/Science Photo Library.

**direct and indirect
life cycles: vectors**

Cutting across the distinction between micro- and macroparasites, parasites can also be subdivided into those that are transmitted directly from host to host and those that require a vector or intermediate host for transmission and therefore have an indirect life cycle. The term ‘vector’ signifies an animal carrying a parasite from

host to host, and some vectors play no other role than as a carrier; but many vectors are also intermediate hosts within which the parasite grows and/or multiplies. Indeed, parasites with indirect life cycles may elude the simple micro/macro distinction. For example, schistosome parasites spend part of their life cycle in a snail and part in a vertebrate (in some cases a human). In the snail, the parasite multiplies and so behaves as

a microparasite, but in an infected human the parasite grows and produces eggs but does not itself multiply, and so behaves as a macroparasite.

12.2.1 Microparasites

Probably the most obvious microparasites are the bacteria and viruses that infect animals (such as the measles virus and the typhoid bacterium) and plants (e.g. the yellow net viruses of beet and tomato and the bacterial crown gall disease). The other major group of microparasites affecting animals is the protozoa (e.g. the trypanosomes that cause sleeping sickness and the *Plasmodium* species that cause malaria). In plant hosts some of the simpler fungi behave as microparasites.

The transmission of a microparasite from one host to another can be in some cases almost instantaneous, as in venereal disease and the short-lived infective agents carried in the water droplets of coughs and sneezes (influenza, measles, etc.). In other species the parasite may spend an extended dormant period 'waiting' for its new host. This is the case with the ingestion of food or water contaminated with the protozoan *Entamoeba histolytica*, which causes amoebic dysentery, and with the plant parasite *Plasmodiophora brassicae*, which causes 'club-root disease' of crucifers.

Alternatively, a microparasite may depend on a vector for its spread. The two most economically important groups of vector-transmitted protozoan parasites of animals are the trypanosomes, transmitted by various vectors including tsetse flies (*Glossina* spp.) and causing sleeping sickness in humans and nagana in domesticated (and wild) mammals, and the various species of *Plasmodium*, transmitted by anopheline mosquitoes and causing malaria. In both these cases, the flies act also as intermediate hosts, i.e. the parasite multiplies within them.

Many plant viruses are transmitted by aphids. In some 'non-persistent' species (e.g. cauliflower mosaic virus), the virus is only viable in the vector for 1 h or so and is often borne only on the aphid's mouthparts. In other 'circulative' species (e.g. lettuce necrotic yellow virus), the virus passes from the aphid's gut to its circulatory system and thence to its salivary glands. Here, there is a latent period before the vector becomes infective, but it then remains infective for an extended period. Finally, there are 'propagative' viruses (e.g. the potato leaf roll virus) that multiply within the aphid. Nematode worms are also widespread vectors of plant viruses.

12.2.2 Macroparasites

The parasitic helminth worms are major macroparasites of animals. The intestinal nematodes of humans, for example, all of which are transmitted directly, are perhaps the most important human intestinal parasites, both in terms of the number of

people infected and their potential for causing ill health. There are also many types of medically important animal macroparasites with indirect life cycles. For example, the tapeworms are intestinal parasites as adults, absorbing host nutrients directly across their body wall and proliferating eggs that are voided in the host's feces. The larval stages then proceed through one or two intermediate hosts before the definitive host (in these cases, the human) is reinfected. The schistosomes, as we have seen, infect snails and vertebrates alternately. Human schistosomiasis (bilharzia) affects the gut wall where eggs become lodged, and also affects the blood vessels of the liver and lungs when eggs become trapped there too. Filarial nematodes are another group of long-lived parasites of humans that all require a period of larval development in a blood-sucking insect. One, *Wuchereria bancrofti*, does its damage (Bancroftian filariasis) by the accumulation of adults in the lymphatic system (classically, but only rarely, leading to elephantiasis). Larvae (microfilariae) are released into the blood and are ingested by mosquitoes, which also transmit more developed, infective larvae back into the host. Another filarial nematode, *Onchocerca volvulus*, which causes 'river blindness', is transmitted by adult blackflies (the larvae of which live in rivers, hence the name of the disease). Here, though, it is the microfilariae that do the major damage when they are released into the skin tissue and reach the eyes.

In addition, there are lice, fleas, ticks and mites and some fungi that attack animals. Lice spend all stages of their life cycle on their host (either a mammal or a bird), and transmission is usually by direct physical contact between host individuals, often between mother and offspring. Fleas, by contrast, lay their eggs and spend their larval lives in the 'home' (usually the nest) of their host (again, a mammal or a bird). The emerging adult then actively locates a new host individual, often jumping and walking considerable distances in order to do so.

Plant macroparasites include the higher fungi that give rise to the mildews, rusts and smuts, as well as the gall-forming and mining insects, and some flowering plants that are themselves parasitic on other plants.

Direct transmission is common amongst the fungal macroparasites of plants. For example, in the development of mildew on a crop of wheat, infection involves contact between a spore (usually wind dispersed) and a leaf surface, followed by penetration of the fungus into or between the host cells, where it begins to grow, eventually becoming apparent as a lesion of altered host tissue. This phase of invasion and colonization precedes an infective stage when the lesion matures and starts to produce spores.

Indirect transmission of plant macroparasites via an intermediate host is common amongst the rust fungi. For example, in black stem rust, infection is transmitted from an annual grass host (especially the cultivated cereals such as wheat) to the barberry shrub (*Berberis vulgaris*) and from the barberry back to wheat. Infections on the cereal are polycyclic, i.e. within a season spores may infect and form lesions that release spores that infect further cereal plants. It is this phase of intense



Figure 12.2 A cuckoo in the nest.
Reproduced by permission of
FLPA/Martin B. Withers.

multiplication by the parasite that is responsible for epidemic outbreaks of disease. On the other hand, the barberry is a long-lived shrub and the rust is persistent within it. Infected barberry plants may therefore serve as persistent foci for the spread of the rust into cereal crops.

**holo- and
hemiparasitic plants**

Plants in a number of families have become specialized as parasites on other flowering plants. These are of two quite distinct types. *Holoparasites*, such as dodder (*Cuscuta* spp.), lack chlorophyll and are wholly dependent on the host plant for their supply of water, nutrients and fixed carbon. *Hemiparasites*, on the other hand, such as the mistletoes (*Phoradendron* spp.), are photosynthetic but have a poorly developed root system of their own, or none at all. They form connections with the roots or stems of other species and draw most or all of their water and mineral nutrients from the host.

12.2.3 Brood and social parasitism

At first sight the presence of a section about cuckoos might seem out of place here. Mostly a host and its parasite come from very distant systematic groups (mammals and bacteria, fish and tapeworms, plants and viruses). In contrast, *brood parasitism* usually occurs between quite closely related species and even between members of the same species. Yet the phenomenon falls clearly within the definition of parasitism (a brood parasite ‘obtains its nutrients from one or a few host individuals, normally causing harm but not causing death immediately’). Brood parasitism is well developed in social insects (sometimes then called social parasitism), where the parasites use workers of

another, usually very closely related species to rear their progeny (Choudhary *et al.*, 1994). The phenomenon is best known, however, amongst birds.

Bird brood parasites lay their eggs in the nests of other birds (Figure 12.2), which then incubate and rear them. They usually depress the nesting success

**the ecological
importance of brood
parasitic birds**

of the host. Amongst ducks, *intraspecific* brood parasitism appears to be most common. Most brood parasitism, however, is *interspecific*. About 1% of all bird species are brood parasites – including about 50% of the species of cuckoos, two genera of finches, five cowbirds and a duck (Payne, 1977). They usually lay only a single egg in the host’s nest and may adjust the host’s clutch size by removing one of its eggs. The developing parasite may evict the host’s eggs or nestlings and harm any survivors by monopolizing parental care. There is therefore the potential for brood parasites to have profound effects on the population dynamics of the host species. However, the frequency of parasitized nests is usually very low (less than 3%), and some time ago Lack (1963) concluded that ‘the cuckoo is an almost negligible cause of egg and nestling losses amongst English breeding birds’. None the less, some impression of the potential importance of brood parasites is apparent from the fact that magpies (*Pica pica*) in populations that coexist with great spotted cuckoos (*Clamator glandarius*) in Europe invest their reproductive effort into laying significantly larger clutches of eggs than those that live free of brood parasitism (Soler *et al.*, 2001) – but those eggs are smaller in compensation. The presumption that this is an evolutionary response to the losses they suffer due to the cuckoos is supported by the fact that magpies that lay larger parasitized clutches do indeed have a higher probability of successfully raising at least some of their own offspring.

host-specific
polymorphisms: *gentes*

Highly host-specific, polymorphic relationships have evolved among brood parasites. For instance, the cuckoo *Cuculus canorum* parasitizes many different host species, but there are different strains ('*gentes*') within the cuckoo species. Individual females of one strain favor just one host species and lay eggs that match quite closely the color and markings of the eggs of the preferred host. Thus, amongst cuckoo females there is marked differentiation between strains in their mitochondrial DNA, which is passed only from female to female, but not at 'microsatellite' loci within the nuclear DNA, which contains material from the male parents, who do not restrict matings to within their own strain (Gibbs *et al.*, 2000). It has long been suggested (Punnett, 1933) that this is possible because the genes controlling egg patterning are situated on the W chromosome, carried only by females. (In birds, unlike mammals, the females are the heterogametic sex.) This has now been established – though in great tits, *Parus major*, rather than in a species of brood parasite (Gosler *et al.*, 2000). Females produce eggs that resemble those of their mothers and maternal grandmothers (from whom they inherit their W chromosome) but not those of their paternal grandmothers. Of course, if female cuckoos lay eggs that look like those of the species with which they were reared, it is also necessary for them to lay their eggs, inevitably or at least preferentially, in the nests of that species. This is most likely to be the result of early 'imprinting' (i.e. a learned preference) within the nest (Teuschl *et al.*, 1998).

12.3 Hosts as habitats

The essential difference between the ecology of parasites and that of free-living organisms is that the habitats of parasites are themselves alive. A living habitat is capable of growth (in numbers and/or size); it is potentially reactive, i.e. it can respond *actively* to the presence of a parasite by changing its nature, developing immune reactions to the parasite, digesting it, isolating or imprisoning it; it is able to evolve; and in the case of many animal parasites, it is mobile and has patterns of movement that dramatically affect dispersal (transmission) from one habitable host to another.

12.3.1 Biotrophic and necrotrophic parasites

The most obvious response of a host to a parasite is for the whole host to die. Indeed, we can draw a distinction between parasites that kill and then continue life on the dead host (*necrotrophic parasites*) and those for which the host must be alive (*biotrophic parasites*). Necrotrophic parasites blur the tidy distinctions between parasites, predators and saprotrophs (see Section 11.1). Insofar as host death is often inevitable and sometimes quite rapid,

necrotrophic parasites are really predators, and once the host is dead they are saprotrophs. But for as long as the host is alive, necroparasites share many features with other types of parasite.

For a biotrophic parasite, the death of its host spells the end of its active life. Most parasites are biotrophic. *Lucilia cuprina*, the blowfly of sheep, however, is a necroparasite on an animal host. The fly lays eggs on the living host and the larvae (maggots) eat into its flesh and may kill it. The maggots continue to exploit the carcass after death but they are now detritivores rather than either parasites or predators. Necroparasites on plants include many that attack the vulnerable seedling stage and cause symptoms known as 'damping-off' of seedlings. *Botrytis fabi* is a typical fungal necroparasite of plants. It develops in the leaves of the bean *Vicia faba*, and the cells are killed, usually in advance of penetration. Spots and blotches of dead tissues form on the leaves and the pods. The fungus continues to develop as a decomposer, and spores are formed and then dispersed from the dead tissue, but not while the host tissue is still alive.

Most necroparasites can therefore be regarded as pioneer saprotrophs. They are one jump ahead of competitors because they can kill the host (or its parts) and so gain first access to the resources of its dead body. The response of the host to necroparasites is never very subtle. Amongst plant hosts, the most common response is to shed the infected leaves, or to form specialized barriers that isolate the infection. Potatoes, for example, form corky scabs on the tuber surface that isolate infections by *Actinomyces scabies*.

necroparasites:
pioneer saprotrophs

12.3.2 Host specificity: host ranges and zoonoses

We saw in the chapters on the interactions between predators and their prey that there is often a high degree of specialization of a particular predator species on a particular species of prey (monophagy). The specialization of parasites on one or a restricted range of host species is even more striking. For any species of parasite (be it tapeworm, virus, protozoan or fungus) the potential hosts are a tiny subset of the available flora and fauna. The overwhelming majority of other organisms are quite unable to serve as hosts: often, we do not know why.

There are, though, some patterns to this specificity. It seems, for example, that the more intimate a parasite's association with a particular host individual, the more likely it is to be restricted to a particular species of host. Thus, for example, most species of bird lice, which spend their entire lives on one host, exploit only one host species, whereas louse flies, which move actively from one host individual to another, can use several species of host (Table 12.1).

The delineation of a parasite's host range, however, is not always as straightforward as one might imagine.

natural and
accidental hosts

Table 12.1 Specialization in ectoparasites that feed on birds and mammals. (After Price, 1980.)

Scientific name	Common name and lifestyle	Number of species	Percentage of species restricted to:		
			1 host	2 or 3 hosts	More than 3 hosts
Philopteridae	Bird lice (spend whole life on one host)	122	87	11	2
Streblidae	Blood-sucking flies (parasitize bats)	135	56	35	9
Oestridae	Botflies (females fly between hosts)	53	49	26	25
Hystrichopsyllidae	Fleas (jump between hosts)	172	37	29	34
Hippoboscidae	Louse flies (are highly mobile)	46	17	24	59

Species outside the host range are relatively easily characterized: the parasite cannot establish an infection within them. But for those inside the host range, the response may range from a serious pathology and certain death to an infection with no overt symptoms. What is more, it is often the ‘natural’ host of a parasite, i.e. the one with which it has coevolved, in which infection is asymptomatic. It is often ‘accidental’ hosts in which infection gives rise to a frequently fatal pathology. (‘Accidental’ is an appropriate word here, since these are often dead-end hosts, that die too quickly to pass on the infection, within which the pathogen cannot therefore evolve – and to which it cannot therefore be adapted.)

plague: a zoonotic infection with humans as accidental hosts

These issues take on not just parasitological but also medical importance in the case of *zoonotic infections*: infections that circulate naturally, and have coevolved, in one or more species of wildlife but also have a pathological effect on humans. A good example is bubonic and pneumonic plague: the human diseases caused by the bacterium *Yersinia pestis*. *Y. pestis* circulates naturally within populations of a number of species of wild rodent: for example, in the great gerbil, *Rhombomys opimus*, in the deserts of Central Asia, and probably in populations of kangaroo rats, *Dipodomys* spp., in similar habitats in southwestern USA. (Remarkably, little is known about the ecology of *Y. pestis* in the USA, despite its widespread nature and potential threat (see Biggins & Kosoy, 2001).) In these species, there are few if any symptoms in most cases of infection. There are, however, other species where *Y. pestis* infection is devastating. Some of these are closely related to the natural hosts. In the USA, populations of prairie dogs, *Cynomys* spp., also rodents, are regularly annihilated by epidemics of plague, and the disease is an important conservation issue. But there are also other species, only very distantly related to the natural hosts, where untreated plague is usually, and rapidly, fatal. Amongst these are humans. Why such a pattern of differential virulence so often occurs – low virulence in the coevolved host, high virulence in some unrelated hosts, but unable even to cause an infection in others – is an important unanswered question in host–pathogen biology. The issue of host–pathogen coevolution is taken up again in Section 12.8.

12.3.4 Habitat specificity within hosts

Most parasites are also specialized to live only in particular parts of their host. Malarial parasites live in the red blood cells of vertebrates. *Theileria* parasites of cattle, sheep and goats live in the lymphocytes of the mammal, and in the epithelial cells, and later in the salivary gland cells, of the tick that is the disease vector, and so on.

By transplanting parasites experimentally from one part of the host’s body to another, it can be shown that many home in on target habitats.

parasites may search for habitats within their hosts

When nematode worms (*Nippostrongylus brasiliensis*) were transplanted from the jejunum into the anterior and posterior parts of the small intestine of rats, they migrated back to their original habitat (Alphey, 1970). In other cases, habitat search may involve growth rather than bodily movement. For instance, loose smut of wheat, the fungus *Ustilago tritici*, infects the exposed stigmas of wheat flowers and then grows as an extending filamentous system into the young embryo. Growth continues in the seedling, and the fungus mycelium keeps pace with the growth of the shoot. Ultimately, the fungus grows rapidly into the developing flowers and converts them into masses of spores.

12.3.5 Hosts as reactive environments: resistance, recovery and immunity

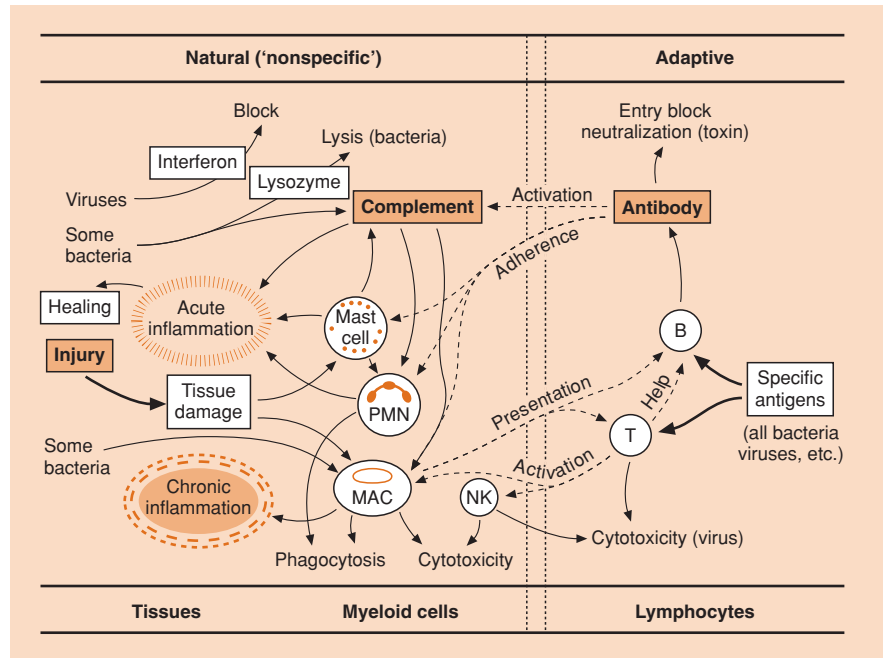
Any reaction by an organism to the presence of another depends on it

invertebrates

recognizing a difference between what is ‘self’ and what is ‘not self’. In invertebrates, populations of phagocytic cells are responsible for much of a host’s response to invaders, even to inanimate particles. In insects, hemocytes (cells in the hemolymph) isolate infective material by a variety of routes, especially encapsulation – responses that are accompanied by the production of a number of soluble compounds in the humoral system that recognize and respond to nonself material, some of which also operate at the midgut barrier in the absence of hemocytes (Siva-Jothy *et al.*, 2001).

Figure 12.3 The immune response.

The mechanisms mediating resistance to infection can be divided into ‘natural’ or ‘nonspecific’ (left) and ‘adaptive’ (right), each composed of both cellular elements (lower half) and humoral elements (i.e. free in the serum or body fluids; upper half). The adaptive response begins when the immune system is stimulated by an antigen that is taken up and processed by a macrophage (MAC). The antigen is a part of the parasite, such as a surface molecule. The processed antigen is presented to T and B lymphocytes. T lymphocytes respond by stimulating various clones of cells, some of which are cytotoxic (NK, natural killer cells), as others stimulate B lymphocytes to produce antibodies. The parasite that bears the antigen can now be attacked in a variety of ways. PMN, polymorphonuclear neutrophil. (After Playfair, 1996.)



vertebrates: the immune response

In vertebrates there is also a phagocytic response to material that is not self, but their armory is considerably extended by a much more elaborate process: the immune response (Figure 12.3). For the ecology of parasites, an immune response has two vital features: (i) it may enable a host to recover from an infection; and (ii) it can give a once-infected host a ‘memory’ that changes its reaction if the parasite strikes again, i.e. the host has become immune to reinfection. In mammals, the transmission of immunoglobulins to the offspring can sometimes even extend protection to the next generation.

For most viral and bacterial infections of vertebrates, the colonization of the host is a brief and transient episode in the host’s life. The parasites multiply within the host and elicit a strong immunological response. By contrast, the immune responses elicited by many of the macroparasites and protozoan microparasites tend to be weaker. The infections themselves, therefore, tend to be persistent, and hosts may be subject to repeated reinfection.

contrasting responses to micro- and macroparasites

Indeed, responses to microparasites and helminths seem often to be dominated by different pathways within the immune system (MacDonald *et al.*, 2002), and these pathways can down-regulate each other: helminth infection may therefore increase the likelihood of microparasitic infection and vice versa (Behnke *et al.*, 2001). Thus, for example, successful treatment of worm infections in patients that were also infected with HIV led to a significant drop in their HIV viral load (Wolday *et al.*, 2002).

plants

The modular structure of plants, the presence of cell walls and the absence of a true circulating system (such as blood or lymph) all make any form of immunological response an inefficient protection. There is no migratory population of phagocytes in plants that can be mobilized to deal with invaders. There is, however, growing evidence that higher plants possess complex systems of defense against parasites. These defenses may be constitutive – physical or biological barriers against invading organisms that are present whether the parasite is present or not – or inducible, arising in response to pathogenic attack (Ryan & Jagendorf, 1995; Ryan *et al.*, 1995). After a plant has survived a pathogenic attack, ‘systematic acquired resistance’ to subsequent attacks may be elicited from the host. For example, tobacco plants infected on one leaf with tobacco mosaic virus can produce local lesions that restrict the virus infection locally, but the plants then also become resistant to new infections not only by the same virus but to other parasites as well. In some cases the process involves the production of ‘elicitors’, which have been purified and shown to induce vigorous defense responses by the host (Yu, 1995).

the costliness of host defense

Central to our understanding of all host defensive responses to parasites is the belief that these responses are costly – that energy and material invested in the response must be diverted away from other important bodily functions – and that there must therefore be a trade-off between the response and other aspects of the life history: the more that is invested in one, the less can be invested

Table 12.2 Estimated energetic costs (percentage increase in resting metabolic rate relative to controls) made by various vertebrate hosts when mounting an immune response to a range of ‘challenges’ that induce such a response. (After Lochmiller & Derenberg, 2000.)

Species	Immune challenge	Cost (%)
Human	Sepsis	30
	Sepsis and injury	57
	Typhoid vaccination	16
Laboratory rat	Interleukin-1 infusion	18
	Inflammation	28
Laboratory mouse	Keyhole limpet hemocyanin injection	30
Sheep	Endotoxin	10–49

in the others. Evidence for this in vertebrates is reviewed by Lochmiller and Derenberg (2000), who illustrate, for example, the energetic price (in terms of an increase in resting metabolic rate) paid by a number of vertebrates when mounting an immune response (Table 12.2).

12.3.6 The consequences of host reaction: *S-I-R*

The variations in mechanisms used by different types of organism to fight infection are clearly interesting and important to parasitologists, medics and veterinarians. They are also important to ecologists working on particular systems, where an understanding of the overall biology is essential. But from the perspective of an ecological overview, the *consequences* for the hosts of these responses are more important, both at the whole organism and the population levels. First, these responses determine where individuals are on the spectrum from ‘wholly susceptible’ to ‘wholly resistant’ to infection – and if they become infected, where they are on the spectrum from being killed by infection to being asymptomatic. Second, in the case of vertebrates, the responses determine whether an individual still expresses a naive susceptibility or has acquired an immunity to infection.

These individual differences then determine, for a population, the structure of that population in terms of the numbers of individuals in the different classes. Many mathematical models of host–pathogen dynamics, for example, are referred to as *S-I-R* models, because they follow the changing numbers of susceptible, infectious and recovered (and immune) individuals in the population. The variations at the population level are then crucial in molding the features at the heart of ecology: the distributions and abundances of the organism concerned. We return to these

questions of epidemic behavior in Section 12.4.2 and thereafter in this chapter.

12.3.7 Parasite-induced changes in growth and behavior

Some parasites induce a new *programed* change in the development of the host. The agromyzid flies and cecidomyid and cynipid wasps that form galls on higher plants are remarkable examples. The insects lay eggs in host tissue, which responds by renewed growth. The galls that are produced are the result of a morphogenetic response that is quite different from any structure that the plant normally produces. Just the presence, for a time, of the parasite egg may be sufficient to start the host tissue into a morphogenetic sequence that can continue even if the developing larva is removed. Amongst the gall-formers that attack oaks (*Quercus* spp.), each elicits a unique morphogenetic response from the host (Figure 12.4).

Fungal and nematode parasites of plants can also induce morphogenetic responses, such as enormous cell enlargement and the formation of nodules and other ‘deformations’. After infection by the bacterium *Agrobacterium tumefaciens*, gall tissue can be recovered from the host plant that lacks the parasite but has now been set in its new morphogenetic pattern of behavior; it continues to produce gall tissue. In this case, the parasite has induced a genetic transformation of the host cells. Some parasitic fungi also ‘take control’ of their host plant and castrate or sterilize it. The fungus *Epichloe typhina*, which parasitizes grasses, prevents them from flowering and setting seed – the grass remains a vegetatively vigorous eunuch, leaving descendant parasites but no descendants of its own.

Most of the responses of modular organisms to parasites (and indeed other environmental stimuli) involve changes in growth and form, but in unitary organisms the response of hosts to infection more often involves a change in behavior: this often increases the chance of transmission of the parasite. In worm-infected hosts, irritation of the anus stimulates scratching, and parasite eggs are then carried from the fingers or claws to the mouth. Sometimes, the behavior of infected hosts seems to maximize the chance of the parasite reaching a secondary host or vector. Praying mantises have been observed walking to the edge of a river and apparently throwing themselves in, whereupon, within a minute of entering the water, a gordian worm (*Gordius*) emerges from the anus. This worm is a parasite of terrestrial insects but depends on an aquatic host for part of its life cycle. It seems that an infected host develops a hydrophilia that ensures that the parasite reaches a watery habitat. Suicidal mantises that are rescued will return to the riverbank and throw themselves in again.

galls

(sometimes dramatic)
changes in host
behavior

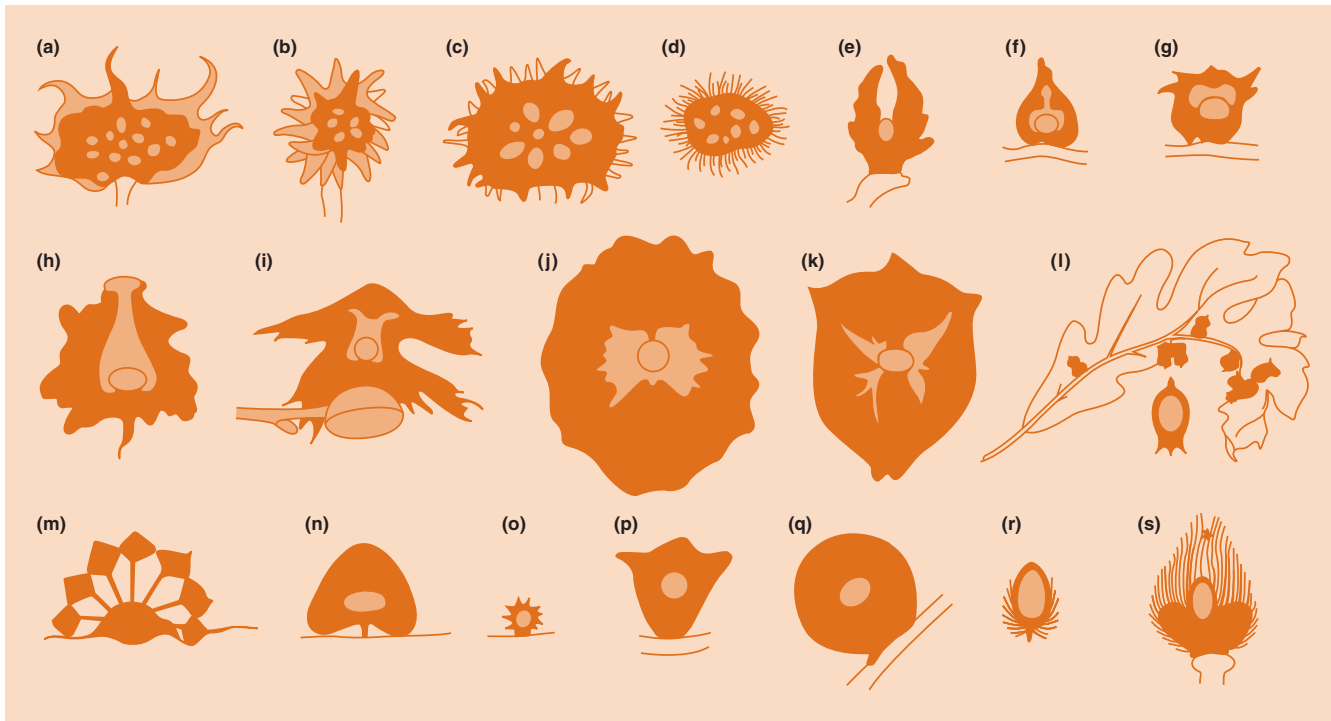


Figure 12.4 Galls formed by wasps of the genus *Andricus* on oaks (*Quercus petraea*, *Q. robur*, *Q. pubescens* or *Q. cerris*). Each figure shows a section through a gall induced by a different species of *Andricus*. The dark colored areas are the gall tissue and the central lighter areas are the cavities containing the insect larva. (From Stone & Cook, 1998.)

12.3.8 Competition within hosts

constant final yield?

Since hosts are the habitat patches for their parasites, it is not surprising that intra- and interspecific competition, observed in other species in other habitats, can also be observed in parasites within their hosts. There are many examples of the fitness of individual parasites decreasing within a host with increasing overall parasite abundance (Figure 12.5a), and of the overall output of parasites from a host reaching a saturation level (Figure 12.5b) reminiscent of the ‘constant final yield’ found in many plant monocultures subject to intraspecific competition (see Section 5.5.1).

competition or the immune response?

However, in vertebrates at least, we need to be cautious in interpreting such results simply as a consequence of intraspecific competition for limited resources, since the intensity of the immune reaction elicited from a host itself typically depends on the abundance of parasites. A rare attempt to disentangle these two effects utilized the availability of mutant rats lacking an effective immune response (Paterson & Viney, 2002). These and normal, control rats were subjected to experimental infection with a nematode, *Strongyloides*

ratti, at a range of doses. Any reduction in parasite fitness with dose in the normal rats could be due to intraspecific competition and/or an immune response that itself increases with dose; but clearly, in the mutant rats only the first of these is possible. In fact, there was no observable response in the mutant rats (Figure 12.6), indicating that at these doses, which were themselves similar to those observed naturally, there was no evidence of intraspecific competition, and that the pattern observed in the normal rats is entirely the result of a density-dependent immune response. Of course, this does not mean that there is never intraspecific competition amongst parasites within hosts, but it does emphasize the particular subtleties that arise when an organism’s habitat is its reactive host.

We know from Chapter 8 that niche differentiation, and especially species having more effect on their own populations than on those of potential competitors, lies at the heart of our understanding of competitor coexistence. We noted earlier that parasites typically specialize on particular sites or tissues within their hosts, suggesting ample opportunity for niche differentiation. And in vertebrates at least, the specificity of the immune response also means that each parasite tends to have its greatest adverse effect on its own population. On the other hand, many parasites do have host tissues and resources in common; and it

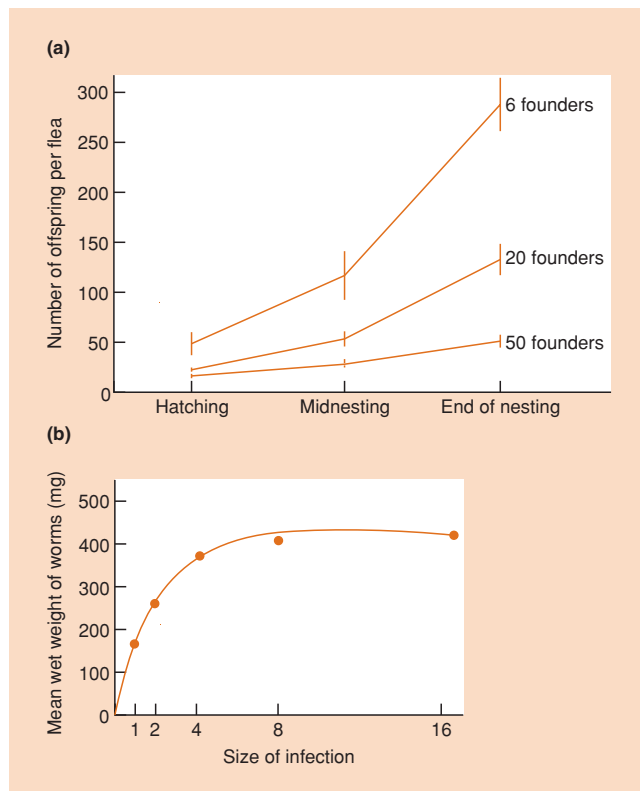


Figure 12.5 Density-dependent responses of parasites within their hosts. (a) The relationship between the number of fleas *Ceratophyllus gallinae* ('founders') added to the nests of blue tits and the number of offspring per flea (mean \pm SE). The greater the density, the lower the reproductive rate of the fleas. This differential increased from an initial assessment at blue tit egg hatching, through to the end of the nestling period. (After Tripet & Richner, 1999.) (b) The mean weight of worms per infected mouse reaches a 'constant final yield' after deliberate infection at a range of levels with the tapeworm *Hymenolepis microstoma*. (After Moss, 1971.)

is easy to see that the presence of one parasite species may make a host less vulnerable to attack by a second species (for example, as a result of inducible responses in plants), or more vulnerable (simply because of the host's weakened state). All in all, it is no surprise that the ecology of parasite competition within hosts is a subject with no shortage of unanswered questions.

None the less, some evidence for interspecific competition amongst parasites comes from a study of two species of nematode, *Howardula aoronymphium* and *Parasitylenchus nearcticus*, that infect the fruit-fly *Drosophila recens* (Perlman & Jaenike, 2001). Of these, *P. nearcticus* is a specialist, being found only in *D. recens*, whereas *H. aoronymphium* is more of a generalist, capable of infect-

interspecific
competition
amongst parasites

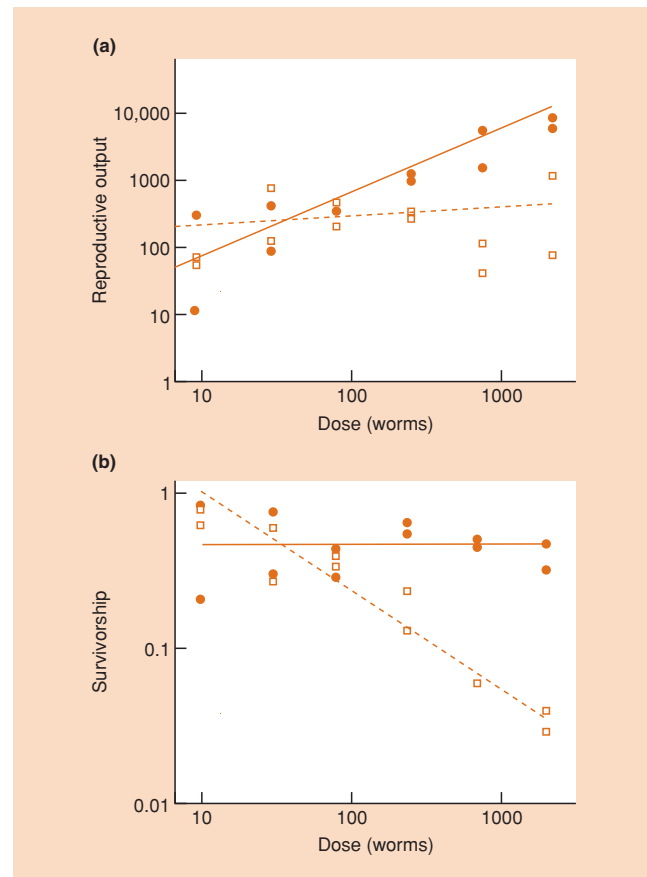


Figure 12.6 Host immune responses are necessary for density dependence in infections of the rat with the nematode *Strongyloides ratti*. (a) Overall reproductive output increases in line with the initial dose in mutant rats without an immune response (\bullet ; slope not significantly different from 1), but with an immune response (\square) it is roughly independent of initial dose, i.e. it is regulated (slope = 0.15, significantly less than 1, $P < 0.001$). (b) Survivorship is independent of the initial dose in mutant rats without an immune response (\bullet ; slope not significantly different from 0), but with an immune response (\square) it declines (slope = -0.62 , significantly less than 0, $P < 0.001$). (After Paterson & Viney, 2002.)

ing a range of *Drosophila* species. In addition, *P. nearcticus* has the more profound effect on its host, typically sterilizing females, whereas *H. aoronymphium* seems to reduce host fecundity by only around 25% (though this itself represents a drastic reduction in host fitness). It is also apparent that whereas *H. aoronymphium* is profoundly affected by *P. nearcticus* when the two coexist within the same host in experimental infections (Figure 12.7a), this effect is not reciprocated (Figure 12.7b). Overall, therefore, competition is strongly asymmetric between the two parasites (as interspecific competition frequently is; see Section 8.3.3): the specialist *P. nearcticus* is both a more powerful exploiter of its host

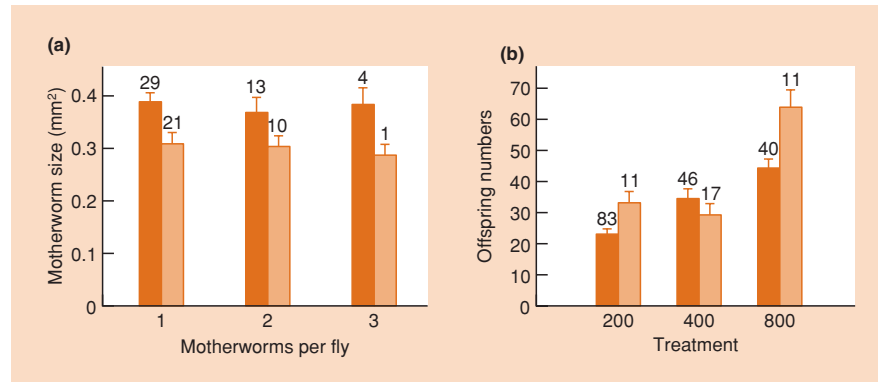


Figure 12.7 (a) Mean size \pm SE (mm^2 , longitudinal section area) of *Howardula aoronymphium* motherworms in 1-week-old hosts, *Drosophila recens*, in single and mixed infections. Size is a good index of fecundity in *H. aoronymphium*. The hosts contained either one, two or three *H. aoronymphium* motherworms, having been reared on a diet contaminated with either *H. aoronymphium* (dark bars) or mixed infections (*H. aoronymphium* and *Parasitylenchus nearcticus*; light bars). Size (fecundity) was consistently lower in mixed infections. (b) Number of *P. nearcticus* offspring (i.e. fecundity) \pm SE, in single (dark bars) and mixed (light bars) infections. Numbers above the bars indicate sample sizes of flies; treatment numbers refer to the numbers of nematodes added to the diet. Fecundity was not reduced in mixed infections. (After Perlman & Jaenike, 2001.)

(reducing it to lower densities through its effect on fecundity) and stronger in interference competition. Coexistence between the species occurs, presumably, because the fly host provides the whole of both the fundamental and the realized niche of *P. nearcticus*, whereas it is only part of the realized niche of *H. aoronymphium*.

12.4 Dispersal (transmission) and dispersion of parasites amongst hosts

12.4.1 Transmission

hosts as islands

Janzen (1968) pointed out that we could usefully think of hosts as islands that are colonized by parasites. By using the same vocabulary, this brought host–parasite relationships into the same arena as MacArthur and Wilson’s (1967) study of island biogeography (see Section 21.5). A human colonized by the malarial parasite is in a sense an inhabited island or patch. The chances of a mosquito vector carrying the parasite from one host to another correspond to the varying distances between different islands. Populations of parasites are thus maintained by the continual colonization of new host patches as old infected patches (hosts) die or become immune to new infection. The whole parasite population is then a ‘metapopulation’ (see Section 6.9), with each host supporting a subpopulation of the whole.

direct and indirect transmission; short- and long-lived agents

Different species of parasite are, of course, transmitted in different ways

between hosts. The most fundamental distinction, perhaps, is that between parasites that are transmitted directly from host to host and those that require a vector or intermediate host for transmission. Amongst the former, we should also distinguish between those where infection is by physical contact between hosts or by a very short-lived infective agent (borne, for example, in coughs and sneezes), and those where hosts are infected by long-lived infective agents (e.g. dormant and persistent spores).

We are largely familiar, through our own experience, with the nature of these distinctions amongst animal pathogens; but essentially the same patterns apply in plants. For example, many soil-borne fungal diseases are spread from one host plant to another by root contacts, or by the growth of the fungus through the soil from a base established on one plant, which gives it the resources from which to attack another. The honey fungus *Armillaria mellea* spreads through the soil as a bootlace-like ‘rhizomorph’ and can infect another host (usually a woody tree or shrub) where it meets their roots. In naturally diverse communities, such spread is relatively slow, but when plants occur as ‘continents’ of continuous interplant contacts, there are greatly increased opportunities for infection to spread. For diseases that are spread by wind, the foci of infection may become established at great distances from the origin; but the rate at which an epidemic develops locally is strongly dependent on the distance between individuals. It is characteristic of wind-dispersed propagules (spores, but also pollen and seeds) that the distribution achieved by dispersal is usually strongly ‘leptokurtic’: a few propagules go a very long way but the majority are deposited close to the origin.

12.4.2 Transmission dynamics

Transmission dynamics are in a very real sense the driving force behind the overall population dynamics of pathogens, but they are often the aspect about which we have least data (compared, say, to the fecundity of parasites or the death rate of infected hosts). We can, none the less, build a picture of the principles behind transmission dynamics (Begon *et al.*, 2002).

The rate of production of new infections in a population, as a result of transmission, depends on the per capita transmission rate (the rate of transmission per susceptible host ‘target’) and the number of susceptible hosts there are (which we can call S). In turn, per capita transmission rate is usually proportional, first, to the contact rate, k , between susceptible hosts and whatever it is that carries the infection. It also depends on the probability, p , that a contact that might transmit infection actually does so. Clearly, this probability depends on the infectiousness of the parasite, the susceptibility of the host, and so on. Putting these three components together we can say:

$$\text{the rate of production of new infections} = k \cdot p \cdot S. \quad (12.1)$$

the contact rate

The details of the contact rate, k , are different for different types of transmission.

- For parasites transmitted directly from host to host, we deal with the rate of contact between infected hosts and susceptible (uninfected) hosts.
- For hosts infected by long-lived infective agents that are isolated from hosts, it is the rate of contact between these and susceptible hosts.
- With vector-transmitted parasites we deal with the contact rate between host and vector (the ‘host-biting rate’), and this goes to determine *two* key transmission rates: from infected hosts to susceptible vectors and from infected vectors to susceptible hosts.

But what is it that determines the per capita contact rate between susceptibles and infecteds? For long-lived infective agents, it is usually assumed that the contact rate is determined essentially by the density of these agents. For direct and vector-borne transmission, however, the contact rate needs to be broken down further into two components. The first is the contact rate between a susceptible individual and *all* other hosts (direct transmission) or all vectors; we can call this c . The second is then the proportion of those hosts or vectors that are infectious; we call this I/N , where I is the number of infecteds and N the total number of hosts (or vectors). Our expanded equation is now:

$$\begin{aligned} \text{the rate of production of new infections} \\ = c \cdot p \cdot S \cdot (I/N). \end{aligned} \quad (12.2)$$

We need to try to understand c and I/N in turn.

12.4.3 Contact rates: density- and frequency-dependent transmission

For most infections, it has often been assumed that the contact rate c increases in proportion to the density of the population, N/A , where A is the area occupied by the population, i.e. the denser the population, the more hosts come into contact with one another (or vectors contact hosts). Assuming for simplicity that A remains constant, the N s in the equation then cancel, all the other constants can be combined into a single constant β , the ‘transmission coefficient’, and the equation becomes:

density-dependent transmission

$$\text{the rate of production of new infections} = \beta \cdot S \cdot I. \quad (12.3)$$

This, unsurprisingly, is known as *density-dependent transmission*.

On the other hand, it has long been asserted that for sexually transmitted diseases, the contact rate is constant: the frequency of sexual contacts is independent of population density. This time the equation becomes:

frequency-dependent transmission

$$\begin{aligned} \text{the rate of production of new infections} \\ = \beta' \cdot S \cdot (I/N), \end{aligned} \quad (12.4)$$

where the transmission coefficient again combines all the other constants but this time acquires a ‘prime’, β' , because the combination of constants is slightly different. This is known as *frequency-dependent transmission*.

Increasingly, however, it has become apparent that the assumed simple correspondence between sexual transmission and frequency dependence on the one hand, and all other types of infection and density dependence on the other, is incorrect. For example, when density and frequency dependence were compared as descriptors of the transmission dynamics of cowpox virus, which is not sexually transmitted, in natural populations of bank voles (*Clethrionomys glareolus*), frequency dependence appeared, if anything, to be superior (Begon *et al.*, 1998). Frequency dependence appears to be a better descriptor than density dependence, too, for a number of (nonsexually transmitted) infections of insects (Fenton *et al.*, 2002). One likely explanation in such cases is that sexual contact is not the only aspect of behavior for which the contact rate varies little with population density: many social contacts, territory defense for instance, may come into the same category.

Secondly, $\beta \cdot S \cdot I$ and $\beta' \cdot S \cdot I/N$ are themselves increasingly recognized (e.g. McCallum *et al.*, 2001) as, at best, benchmarks against which real examples of transmission might be measured, rather than exact descriptors of the dynamics; or perhaps as ends of a spectrum along which real transmission terms could be assembled. For example, fitting the term $\beta S^x I^y$ to the transmission dynamics of granulovirus infection in larvae of the moth *Plodia interpunctella* revealed that the best fit was not to ‘pure’ density-dependent transmission, βSI , but to $\beta' S^{1.12} I^{0.14}$ (Figure 12.8).

ends of a spectrum

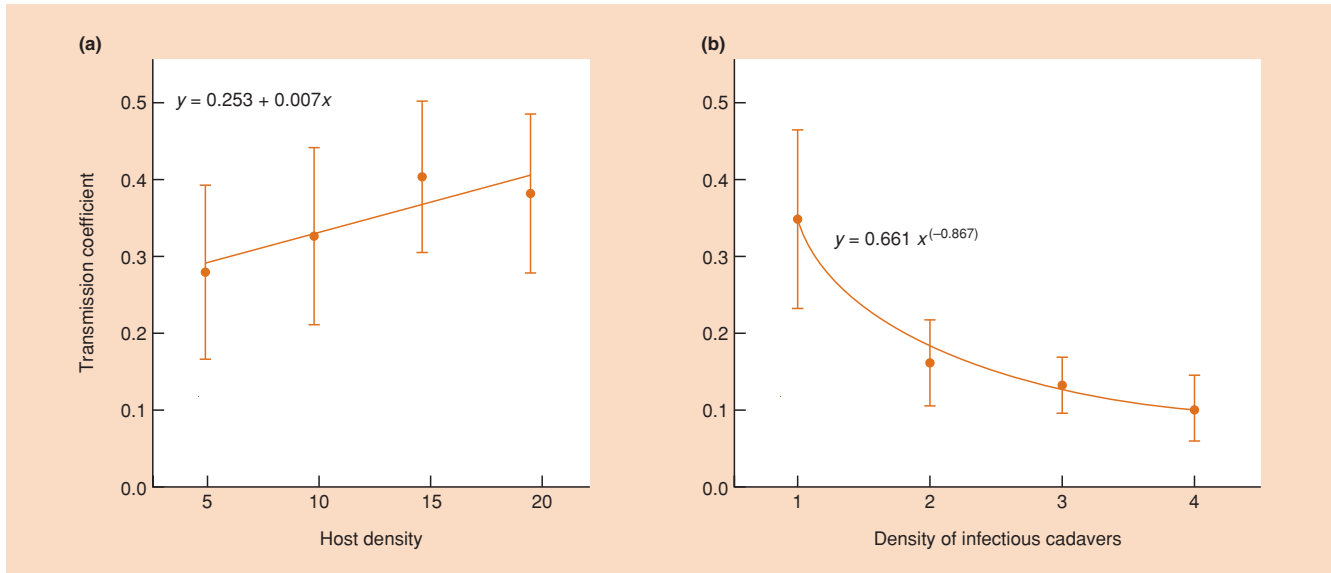


Figure 12.8 Estimating the transmission coefficient at various densities of (a) susceptible hosts and (b) infectious cadavers during the transmission of a granulovirus amongst moths, *Plodia interpunctella*, showed that the coefficient appeared to increase with the former and decrease with the latter. This is contrary to the expectations from density-dependent transmission (an apparently constant coefficient in both cases). (After Knell *et al.*, 1998.)

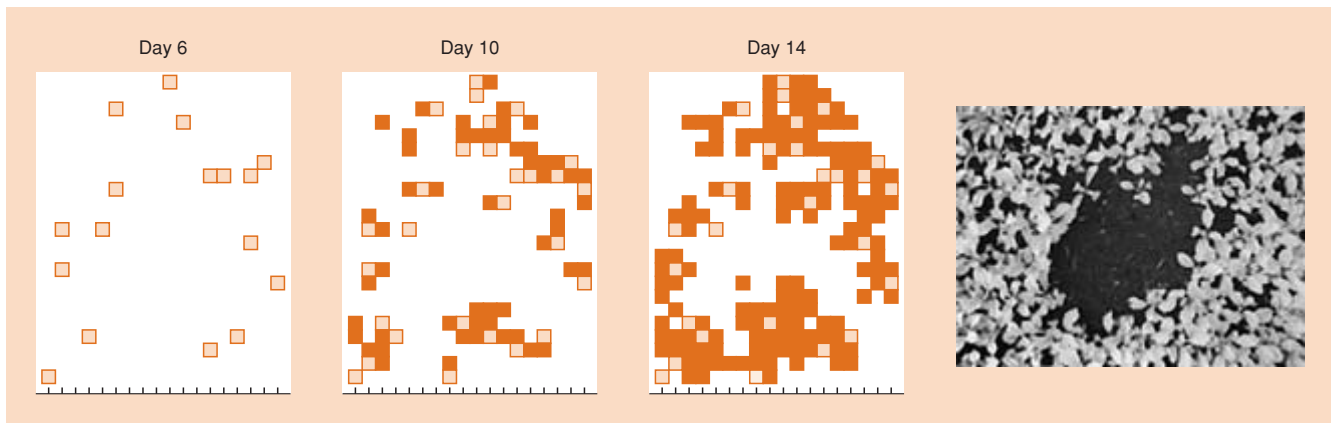


Figure 12.9 The spatial spread of damping-off disease in a population of radish plants, *Raphanus sativus*, caused by the fungus *Rhizoctonia solani*. Following initiation of the disease at isolated plants (light squares), the epidemic spreads rapidly to neighboring plants (dark squares), resulting in patches of damped-off plants (picture on the right). (Courtesy of W. Otten and C.A. Gilligan, Cambridge University.)

In other words, transmission was greater than expected (exponent greater than 1) at higher densities of susceptible hosts, probably because hosts at higher densities were short of food, moved more, and consumed more infectious material. But it was lower than expected (exponent less than 1) at higher densities of infectious host cadavers, probably because of strongly differential susceptibility amongst the hosts, such that the most susceptible become infected even at low cadaver densities, but the least susceptible remain uninfected even as cadaver density increases.

Turning from the contact rate, c , to the I/N term, there has usually been a simplifying assumption that this can

local hot spots

be based on numbers from the whole of a population. In reality, however, transmission typically occurs locally, between nearby individuals. In other words, use of such a term assumes either that all individuals in a population are intermingling freely with one another, or, slightly more realistically, that individuals are distributed approximately evenly across the population, so that all susceptibles are subject to roughly the same probability of a contact being with an infectious individual, I/N . The reality, however, is that there are likely to be hot spots of infection in a population, where I/N is high, and corresponding cool zones. Transmission, therefore, often gives rise to spatial waves of infection passing through a population (e.g. Figure 12.9), rather than simply the

overall rise in infection implied by a global transmission term like βSI . This illustrates a very general point in modeling: that is, the price paid in diminished realism when a complex process is boiled down into a simple term (such as βSI). None the less, as we shall see (and have seen previously in other contexts) without such simple terms to help us, progress in understanding complex processes would be impossible.

12.4.4 Host diversity and the spatial spread of disease

The further that hosts are isolated from one another, the more remote are the chances that a parasite will spread between them. It is perhaps no surprise, then, that the major disease epidemics known amongst plants have occurred in crops that are not islands in a sea of other vegetation, but ‘continents’ – large areas of land occupied by one single species (and often by one single variety of that species). Conversely, the spatial spread of an infection can be slowed down or even stopped by mixtures of susceptible and resistant species or varieties (Figure 12.10). A rather similar effect is described in Section 22.3.1.1, for Lyme disease in the United States, where a variety of host species that are incompetent in transmitting the spirochete pathogen ‘dilute’ transmission between members of the most competent species.

In agricultural practice, resistant cultivars offer a challenge to evolving parasites: mutants that can attack the resistant strain have an immediate gain in fitness. New, disease-resistant crop varieties therefore tend to be widely adopted into commercial practice; but they then often succumb, rather suddenly, to a different race of the pathogen. A new resistant strain of crop is then used, and in due course a new race of pathogen emerges. This ‘boom and bust’ cycle is repeated endlessly and keeps the pathogen in a continually evolving condition, and plant breeders in continual employment. An escape from the cycle can be gained by the deliberate mixing of varieties so that the crop is dominated neither by one virulent race of the pathogen nor by one susceptible form of the crop itself.

the Janzen–Connell effect

In nature there may be a particular risk of disease spreading from perennial plants to seedlings of the same species growing close to them. If this were commonly the case, it could contribute to the species richness of communities by preventing the development of monocultures. This has been called the Janzen–Connell effect. In an especially complete test for the effect, Packer and Clay (2000) showed for black cherry, *Prunus serotina*, trees in a woodland in Indiana, first, that seedlings were indeed less likely to survive close to their parents (Figure 12.11a). Second, they showed that it was something in the soil close to the parents that reduced survival (Figure 12.11b), though this was only apparent at high seedling density, and the effect could be removed by sterilizing the soil. This suggests a pathogen, which high densities of seedlings, close to the parent, amplify and transmit to other seedlings. In fact, dying

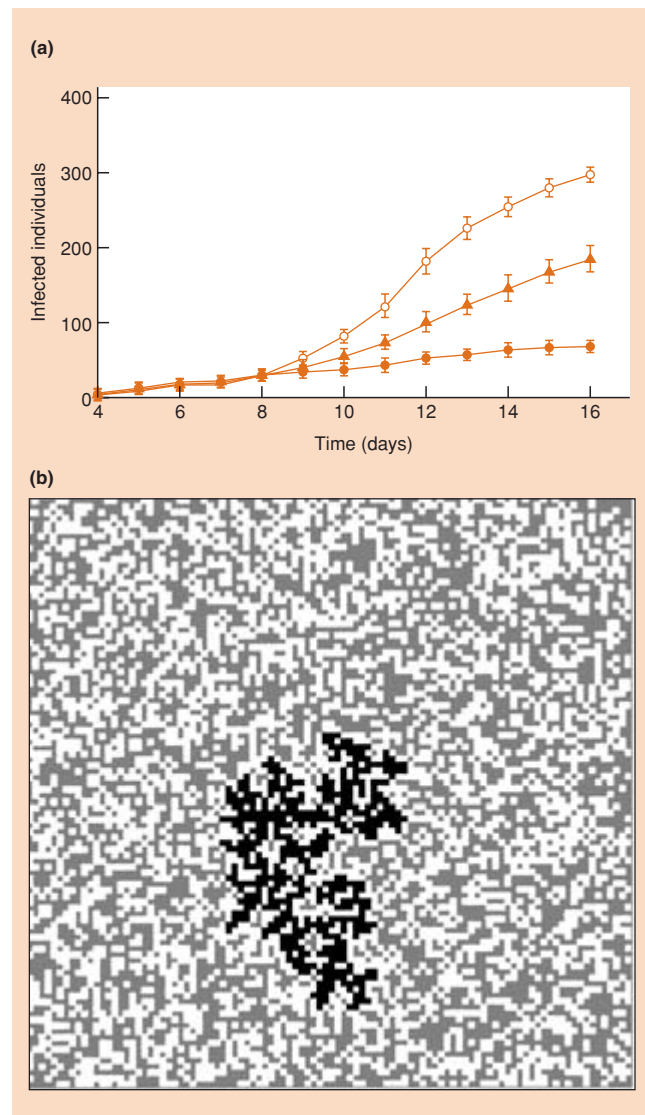


Figure 12.10 The effect of resistant forms in slowing down the spread of damping-off epidemics caused by the fungus *Rhizoctonia solani*. (a) Progress of epidemics in populations following the introduction of *R. solani* into a susceptible population (radish, *Raphanus sativus*: ○), a partially resistant population (mustard, *Sinapsis alba*: ●) or a 50 : 50 mixture of the two (▲). (b) A simulation showing that when 40% of the plants in a population are of a resistant variety, the spread of a damping-off epidemic following its introduction can be prevented. White squares are resistant plants, black squares are infected, and gray squares susceptible. Infection can only be transmitted to an adjacent plant (sharing a ‘side’). Here, the epidemic can spread no further. (Courtesy of W. Otten, J. Ludlam and C.A. Gilligan, Cambridge University.)

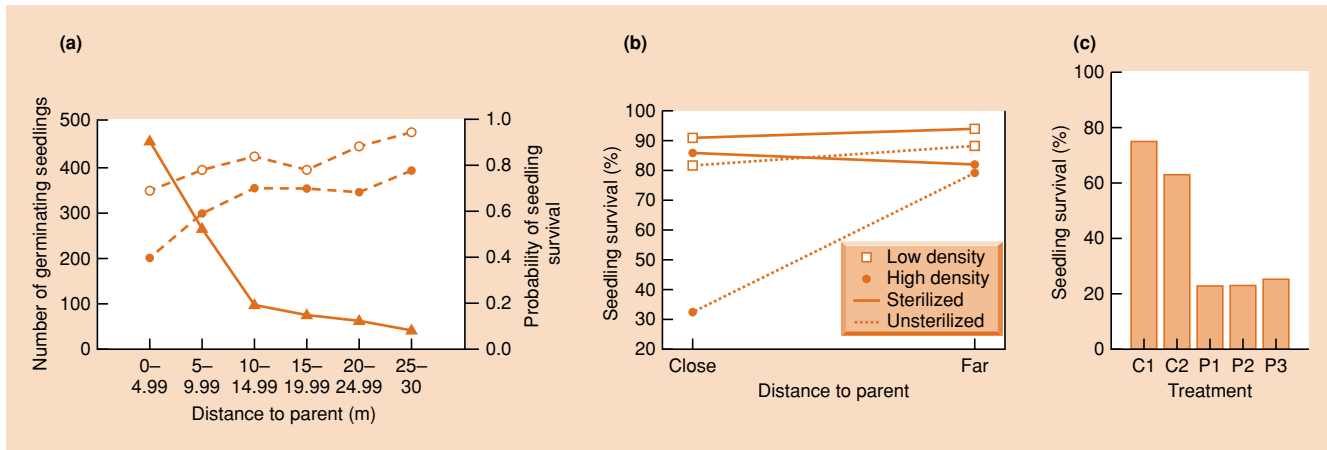


Figure 12.11 (a) The relationship between distance to parent, initial seedling germination (\blacktriangle) and probability of seedling survival over time (dashed lines: \circ , after 4 months; \bullet , after 16 months); $n = 974$ seedlings from beneath six trees. (b) The effect of distance from parent, seedling density and soil sterilization on seedling survival when seedlings were grown in pots containing soil collected close to or far from their parents. In high-density treatments, survival was significantly greater after the soil collected close to the tree was sterilized. ($P < 0.0001$). (c) Seedling survival in control and pathogen inoculation treatments ($n = 40$ per treatment). Control 1, potting mix only; control 2, 5 ml of sterile nutrient-rich fungal growth medium plus potting mix; P1, P2 and P3, three 5 ml replicates of pathogen inoculum plus potting mix. Survival was significantly lower in pathogen treatments compared with controls after 19 days ($X^2 = 13.8$, d.f. = 4, $P < 0.05$). (After Packer & Clay, 2000.)

seedlings were observed with the symptoms of ‘damping off’, and the damping-off fungus, *Pythium* sp., was isolated from dying seedlings and itself caused a significant reduction in seedling survival (Figure 12.11c).

12.4.5 The distribution of parasites within host populations: aggregation

Transmission naturally gives rise to an ever-changing dispersion of parasites within a population of hosts. But if we freeze the frame (or more correctly, carry out a cross-sectional survey of a population at one point in time), then we generate a distribution of parasites within the host population. Such distributions are rarely random. For any particular species of parasite it is usual for many hosts to harbor few or no parasites, and a few hosts to harbor many, i.e. the distributions are usually aggregated or clumped (Figure 12.12).

In such populations, the mean density of parasites (mean number per host) may have little meaning. In a human population in which only one person is infected with anthrax, the mean density of *Bacillus anthracis* is a particularly useless piece of information. A more useful statistic, especially for microparasites, is the *prevalence* of infection: the proportion or percentage of a host population that is infected. On the other hand, infection *may* often vary in severity between individuals and is often clearly related to the number

of parasites that they harbor. The number of parasites in or on a particular host is referred to as the *intensity* of infection. The *mean intensity* of infection is then the mean number of parasites per host in a population (including those hosts that are not infected).

Aggregations of parasites within hosts may arise because individual hosts vary in their susceptibility to infection (whether due to genetic, behavioral or environmental factors), or because individuals vary in their exposure to parasites (Wilson *et al.*, 2002). The latter is especially likely to arise because of the local nature of transmission, and especially when hosts are relatively immobile. Infection then tends to be concentrated, at least initially, close to an original source of infection, and to be absent in individuals in areas that infection has yet to reach, or where it was previously but the hosts have recovered. It is clear, for example, even without explicit data on the distribution of parasites amongst hosts, that the parasites in Figure 12.9, at any one point in time, were aggregated at high intensities around the wave front – but absent ahead of and after it.

12.5 Effects of parasites on the survivorship, growth and fecundity of hosts

According to strict definition, parasites cause harm to their host. But it is not always easy to demonstrate this harm, which may be detectable only at some peculiarly sensitive stage of the host’s life history or under particular circumstances (Toft & Karter, 1990). Indeed, there are examples of ‘parasites’ that feed on a host but

prevalence, intensity
and mean intensity

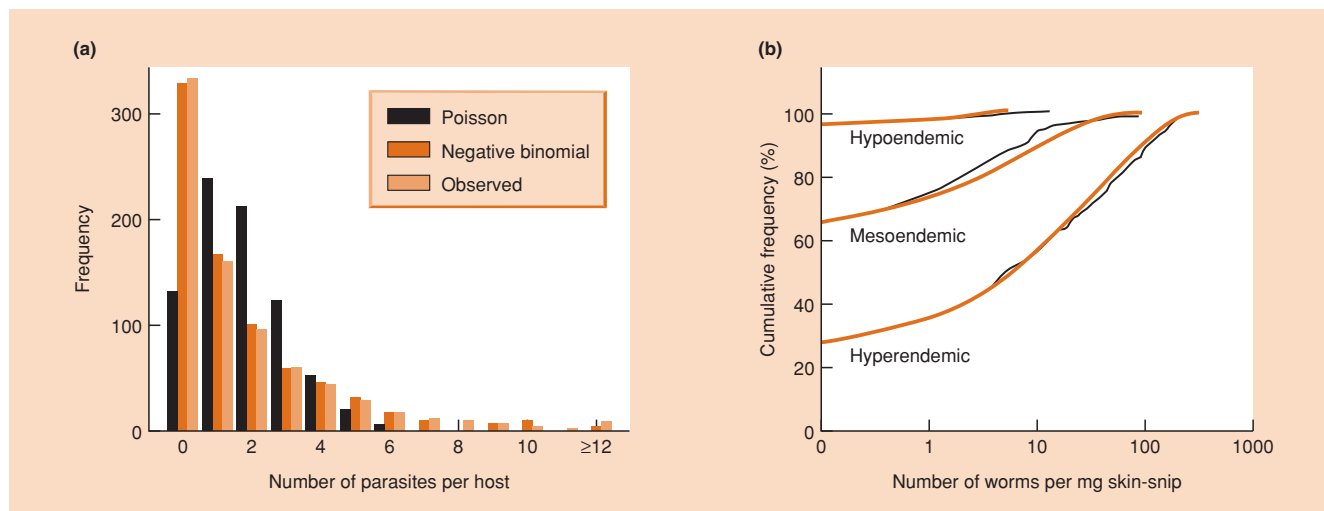


Figure 12.12 Examples of aggregated distributions of parasite numbers per host. (a) Crayfish, *Orconectes rusticus*, infected with the flatworm *Paragonimus kellicotti*. The distribution is significantly different from Poisson (random) ($X^2 = 723$, $P < 0.001$) but conforms well with a ‘negative binomial’, which is good at describing aggregated distributions: $X^2 = 12$, $P \approx 0.4$. (After Stromberg *et al.*, 1978; Shaw & Dobson, 1995.) (b) Distribution of *Onchocerca volvulus* worms, which cause onchocerciasis or ‘river blindness’, in human Yanomami communities in southern Venezuela. Again the distributions, plotted as cumulative frequencies (black lines), conform well to a negative binomial distribution (colored lines), whether the typical intensity of infection is low (hypoendemic), moderate (mesoendemic) or high (hyperendemic). (After Vivas-Martinez *et al.*, 2000.)

appear to do it no harm. For example, in natural populations of Australia’s sleepy lizard, *Tiliqua rugosa*, longevity was either not correlated or was positively associated with their load of ectoparasitic ticks (*Aponomma hydrosauri* and *Amblyomma limbatum*). There was no evidence that the ticks reduced host fitness (Bull & Burzacott, 1993).

There are of course, none the less, examples in which a detrimental effect of a parasite on host fitness has been demonstrated. Table 12.3, for example, shows one particular compilation

of studies in which experimental manipulation of the loads of animal parasites revealed effects on either host fecundity or survival. (And while an effect on fecundity may seem less drastic than one on mortality, this seems less to be the case if one thinks of it as the death of potentially large numbers of offspring.)

On the other hand, the effects of parasites are often more subtle than a simple reduction in survival or fecundity. For example, the pied flycatcher

effects are often subtle . . .

Table 12.3 The impact of various parasites on the fecundity and survival of wild animals, as demonstrated through the experimental manipulation of parasite loads. (After Tompkins & Begon, 1999, where the original references may be found.)

Host	Parasite	Impact
Anderson’s gerbil (<i>Gerbillus andersoni</i>)	<i>Synoternus cleopatrae</i> (flea)	Reduced survival
Barn swallow (<i>Hirundo rustica</i>)	<i>Ornithonyssus bursa</i> (mite)	Reduced fecundity
Cliff swallow (<i>Hirundo pyrrhonota</i>)	<i>Oeciacus vicarius</i> (bug)	Reduced fecundity
European starling (<i>Sturnus vulgaris</i>)	<i>Dermanyssus gallinae</i> (mite)	Reduced fecundity
	<i>Ornithonyssus sylvarius</i> (mite)	Reduced fecundity
Great tit (<i>Parus major</i>)	<i>Ceratophyllus gallinae</i> (flea)	Reduced fecundity
House martin (<i>Delichon urbica</i>)	<i>Oeciacus hirundinis</i> (bug)	Reduced fecundity
Pearly-eyed thrasher (<i>Margarops fuscatus</i>)	<i>Phlebotomus perniciosus</i> (fly)	Reduced fecundity
Purple martin (<i>Progne subis</i>)	<i>Dermanyssus prognepphilus</i> (mite)	Reduced fecundity
Red grouse (<i>Lagopus lagopus</i>)	<i>Trichostrongylus tenuis</i> (nematode)	Reduced fecundity
Snowshoe hare (<i>Lepus americanus</i>)	<i>Obeliscooides cuniculi</i> (nematode)	Reduced survival
Soay sheep (<i>Ovis aries</i>)	<i>Teladorsagia circumcincta</i> (nematode)	Reduced survival

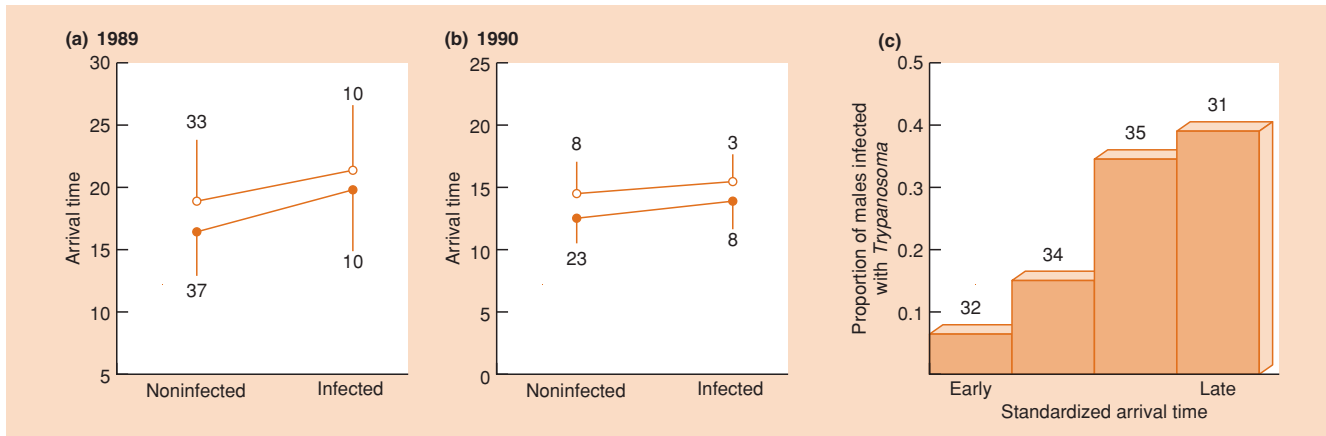


Figure 12.13 The mean date of arrival (1 = May 1) in Finland of male pied flycatchers (*Ficedula hypoleuca*) infected and uninfected with *Trypanosoma*: (a) 1989 and (b) 1990. ●, adult males; ○, yearling males. Sample sizes are indicated near the standard deviation bars. (c) The proportion of males infected with *Trypanosoma* amongst groups of migrants arriving in Finland at different times. (After Rätti *et al.*, 1993.)

(*Ficedula hypoleuca*) migrates from tropical West Africa to Finland to breed, and males that arrive early are particularly successful in finding mates. Males infected with the blood parasite *Trypanosoma* have shorter tails, tend to have shorter wings and arrive in Finland late and so presumably mate less often (Figure 12.13). Another example is provided by lice that feed on the feathers of birds and are commonly regarded as 'benign' parasites, with little or no effects on the fitness of their hosts. However, a long-term comparison of the effects of lice on feral rock doves (*Columba livia*) showed that the lice reduced the thermal protection given by the feathers and, in consequence, heavily infected birds incurred the costs of requiring higher metabolic rates to maintain their body temperatures (Booth *et al.*, 1993) and in the time that the birds spent in preening to keep the lice population under control.

In a similar vein, infection may make hosts more susceptible to predation. For example, postmortem examination of red grouse (*Lagopus lagopus scoticus*) showed that birds killed by predators carried significantly greater burdens of the parasitic nematode *Trichostrongylus tenuis* than the presumably far more random sample of birds that were shot (Hudson *et al.*, 1992a). Alternatively, the effect of parasitism may be to weaken an aggressive competitor and so allow weaker associated species to persist. For example, of two *Anolis* lizards that live on the Caribbean island of St Maarten, *A. gingivinus* is the stronger competitor and appears to exclude *A. wattsi* from most of the island. But the malarial parasite *Plasmodium azurophilum* very commonly affects *A. gingivinus* but rarely affects *A. wattsi*. Wherever the parasite infects *A. gingivinus*, *A. wattsi* is present; wherever the parasite is absent, only *A. gingivinus* occurs (Schall, 1992). Similarly, the holoparasitic plant, dodder (*Cuscuta salina*), which has a strong preference for *Salicornia* in a southern Californian salt marsh, is highly instrumental in determining the outcome of competition between

Salicornia and other plant species within several zones of the marsh (Figure 12.14).

These latter examples make an important point. Parasites often affect their hosts not in isolation, but through an interaction with some other factor: infection may make a host more vulnerable to competition or predation; or competition or shortage of food may make a host more vulnerable to infection or to the effects of infection. This does not mean, however, that the parasites play only a supporting role. Both partners in the interaction may be crucial in determining not only the overall strength of the effect but also which particular hosts are affected.

... affecting an interaction

Organisms that are resistant to parasites avoid the costs of parasitism, but, as with resistance to other natural enemies, resistance itself may carry a cost. This was tested with two cultivars of lettuce (*Lactuca sativa*), resistant or susceptible by virtue of two tightly linked genes to leaf root aphid (*Pemphigus bursarius*) and downy mildew (*Bremia lactucae*). The parasites were controlled by weekly applications of insecticides and fungicides. Resistant forms of lettuce bore fewer axillary buds than susceptibles (Figure 12.15), and this cost of resistance was most marked when the plants were making poor growth because of nutrient deficiency. In nature, hosts must always be caught between the costs of susceptibility and the costs of resistance.

Establishing that parasites have a detrimental effect on host characteristics of demographic importance is a critical first step in establishing that parasites influence the population and community dynamics of their hosts. But it is only a first step. A parasite may increase mortality, directly or indirectly, or decrease fecundity, without this affecting levels or patterns of abundance. The effect may simply be too trivial to have a measurable effect at the population level, or other factors and processes may act in

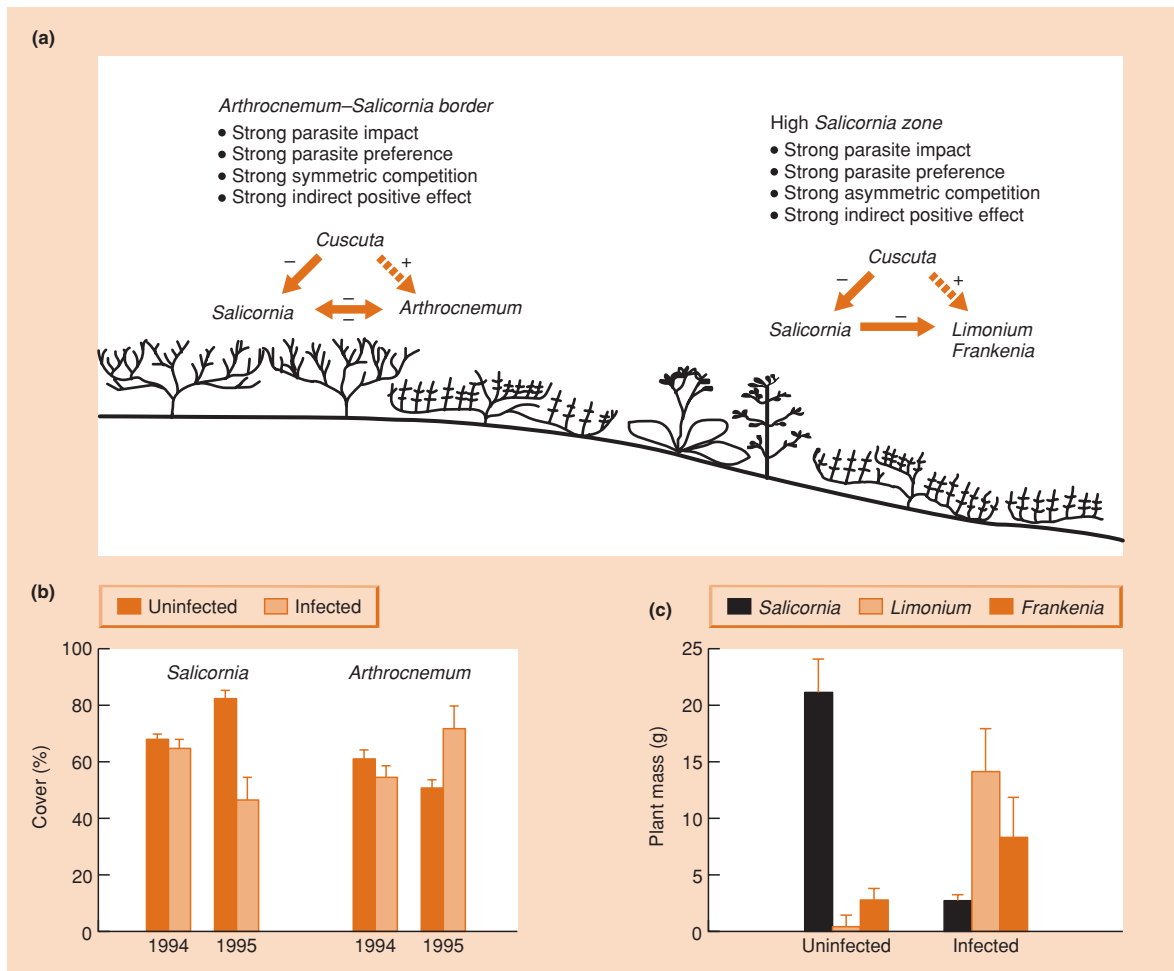


Figure 12.14 The effect of dodder, *Cuscuta salina*, on competition between *Salicornia* and other species in a southern Californian salt marsh. (a) A schematic representation of the main plants in the community in the upper and middle zones of the marsh and the interactions between them (solid lines: direct effects; dashed lines: indirect effects). *Salicornia* (the relatively low-growing plant in the figure) is most attacked by, and most affected by, dodder (which is not itself shown in the figure). When uninfected, *Salicornia* competes strongly and symmetrically with *Arthrocnemum* at the *Arthrocnemum–Salicornia* border, and is a dominant competitor over *Limonium* and *Frankenia* in the middle (high *Salicornia*) zone. However, dodder significantly shifts the competitive balances. (b) Over time, *Salicornia* decreased and *Arthrocnemum* increased in plots infected with dodder. (c) Large patches of dodder suppress *Salicornia* and favor *Limonium* and *Frankenia*. (After Pennings & Callaway, 2002.)

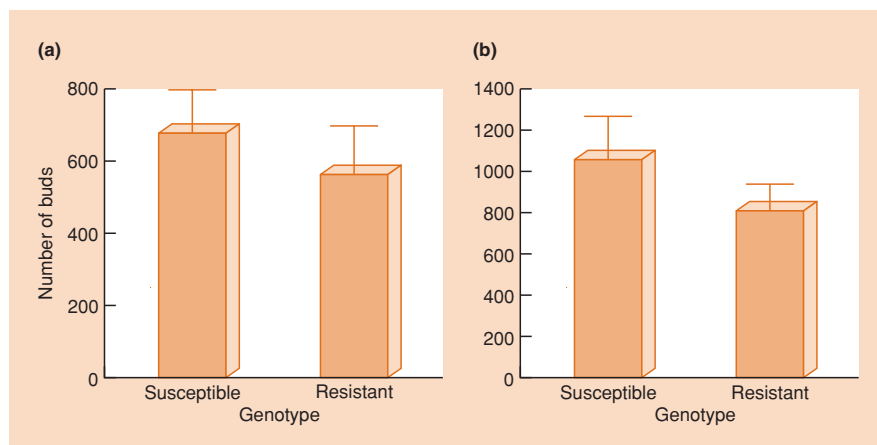


Figure 12.15 The number of buds produced by resistant and susceptible genotypes of two cultivars of lettuce, (a) and (b). Error bars are ± 2 SE. (After Bergelson, 1994.)

a compensatory fashion – for example, loss to parasites may lead to a weakening of density-dependent mortality at a later stage in the life cycle. The effects of rare, devastating epidemics, whether in humans, other animals or plants, are easy to see; but for more typical, endemic parasites and pathogens, moving from the host-individual to the host-population level offers an immense challenge.

12.6 The population dynamics of infection

In principle, the sorts of conclusions that were drawn in Chapter 10 regarding the population dynamics of predator–prey and herbivore–plant interactions can be extended to parasites and hosts. Parasites harm individual hosts, which they use as a resource. The way in which this affects their populations varies with the densities of both parasites and hosts and with the details of the interaction. In particular, infected and uninfected hosts can exhibit compensatory reactions that may greatly reduce the effects on the host population as a whole. Theoretically, a range of outcomes can be predicted: varying degrees of reduction in host-population density, varying levels of parasite prevalence and various fluctuations in abundance.

effects on health or morbidity

With parasites, however, there are particular problems. One difficulty is that parasites often cause a reduction in the ‘health’ or ‘morbidity’ of their host rather than its immediate death, and it is therefore usually difficult to disentangle the effects of the parasites from those of other factors with which they interact (see Section 12.5). Another problem is that even when parasites cause a death, this may not be obvious without a detailed postmortem examination (especially in the case of microparasites). Also, the biologists who describe themselves as parasitologists have in the past tended to study the biology of their chosen parasite without much consideration of the effects on whole host populations; while ecologists have tended to ignore parasites. Plant pathologists and medical and veterinary parasitologists, for their part, generally study parasites with known severe effects that live typically in dense and aggregated populations of hosts, paying little attention to the more typical effects of parasites in populations of ‘wildlife’ hosts. Elucidation of the role of parasites in host-population dynamics is one of the major challenges facing ecology.

Here, we begin by looking at the dynamics of infection within host populations without considering any possible effects on the total abundance of hosts. This ‘epidemiological’ approach (Anderson, 1991) has especially dominated the study of human disease, where total abundance is usually considered to be determined by a whole spectrum of factors and is thus effectively independent of the prevalence of any one infection. Infection only affects the partitioning of this population into susceptible (uninfected), infected and other classes. We then take a more ‘ecological’ approach by considering the effects of parasites on host abundance

in a manner much more akin to conventional predator–prey dynamics.

12.6.1 The basic reproductive rate and the transmission threshold

In all studies of the dynamics of parasite populations or the spread of infection, there are a number of particularly

R_0 , the basic reproductive rate

key concepts. The first is the *basic reproductive rate*, R_0 . For microparasites, because infected hosts are the unit of study, this is defined as the average number of new infections that would arise from a single infectious host introduced into a population of susceptible hosts. For macroparasites, it is the average number of established, reproductively mature offspring produced by a mature parasite throughout its life in a population of uninfected hosts.

The *transmission threshold*, which must be crossed if an infection is to spread, is then given by the condition

the transmission threshold

$R_0 = 1$. An infection will eventually die out for $R_0 < 1$ (each present infection or parasite leads to less than one infection or parasite in the future), but an infection will spread for $R_0 > 1$.

Insights into the dynamics of infection can be gained by considering the various determinants of the basic reproductive rate. We do this in some detail for directly transmitted microparasites, and then deal more briefly with related issues for indirectly transmitted microparasites, and directly and indirectly transmitted macroparasites.

12.6.2 Directly transmitted microparasites: R_0 and the critical population size

For microparasites with direct, density-dependent transmission (see Section 12.4.3), R_0 can be said to increase with: (i) the average period of time over which an infected host remains infectious, L ; (ii) the number of susceptible individuals in the host population, S , because greater numbers offer more opportunities for transmission of the parasite; and (iii) the transmission coefficient, β (see Section 12.4.3). Thus, overall:

$$R_0 = S\beta L. \quad (12.5)$$

Note immediately that by this definition, the greater the number of susceptible hosts, the higher the basic reproductive rate of the infection (Anderson, 1982).

The transmission threshold can now be expressed in terms of a *critical population size*, S_T , where, because $R_0 = 1$ at that threshold:

the critical population size . . .

$$S_T = 1/(\beta L). \quad (12.6)$$

In populations with numbers of susceptibles less than this, the infection will die out ($R_0 < 1$). With numbers greater than this the infection will spread ($R_0 > 1$). (S_T is often referred to as the critical *community size* because it has mostly been applied to human ‘communities’, but this is potentially confusing in a wider ecological context.) These simple considerations allow us to make sense of some very basic patterns in the dynamics of infection (Anderson, 1982; Anderson & May, 1991).

... for different types of parasite

Consider first the kinds of population in which we might expect to find different sorts of infection. If microparasites are highly infectious (large β s), or give rise to long periods of infectiousness (large L s), then they will have relatively high R_0 values even in small populations and will therefore be able to persist there (S_T is small). Conversely, if parasites are of low infectivity or have short periods of infectiousness, they will have relatively small R_0 values and will only be able to persist in large populations. Many protozoan infections of vertebrates, and also some viruses such as herpes, are persistent within individual hosts (large L), often because the immune response to them is either ineffective or short lived. A number of plant diseases, too, like club-root, have very long periods of infectiousness. In each case, the critical population size is therefore small, explaining why they can and do survive endemically even in small host populations.

On the other hand, the immune responses to many other human viral and bacterial infections are powerful enough to ensure that they are only very transient in individual hosts (small

L), and they often induce lasting immunity. Thus, for example, a disease like measles has a critical population size of around 300,000 individuals, and is unlikely to have been of great importance until quite recently in human biology. However, it generated major epidemics in the growing cities of the industrialized world in the 18th and 19th centuries, and in the growing concentrations of population in the developing world in the 20th century. Around 900,000 deaths occur each year from measles infection in the developing world (Walsh, 1983).

12.6.3 Directly transmitted microparasites: the epidemic curve

The value of R_0 itself is also related to the nature of the *epidemic curve* of an infection. This is the time series of new cases following the introduction of the parasite into a population of hosts. Assuming there are sufficient susceptible hosts present for the parasite to invade (i.e. the critical population size, S_T , is exceeded), the initial growth of the epidemic will be rapid as the parasite sweeps through the population of susceptibles. But as these susceptibles either die or recover to immunity, their number, S , will decline, and so too therefore will R_0 (Equation 12.5). Hence, the rate of appearance of new cases will slow down and then decline. And if S falls below S_T and stays there, the infection will disappear – the epidemic will have ended. Two examples of epidemic curves, for Legionnaires’ disease in Spain and for foot-and-mouth disease in the UK, are shown in Figure 12.16.

Not surprisingly, the higher the initial value of R_0 , the more rapid will be the rise in the epidemic curve. But this will also lead

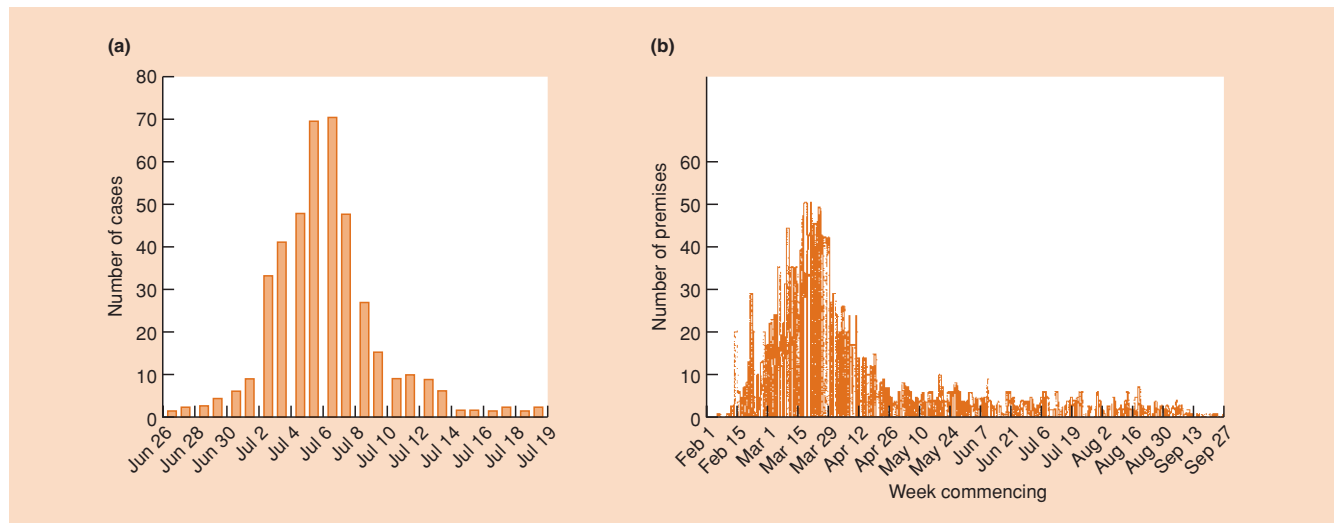
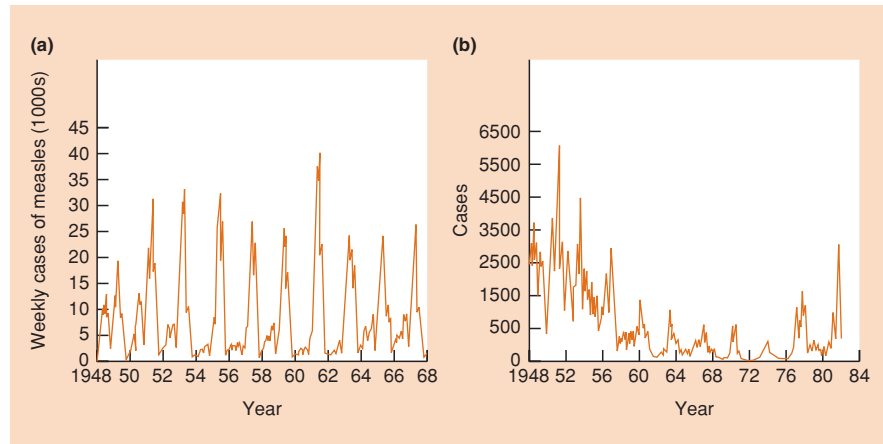


Figure 12.16 (a) An epidemic curve for an outbreak of Legionnaires’ disease in Murcia, a municipality in southeastern Spain, in 2001. (After García-Fulgueiras *et al.*, 2003.) (b) An epidemic curve for an outbreak of foot-and-mouth disease (mostly affecting cattle and sheep) in the United Kingdom in 2001. Infected premises (farms) are shown, since infection was transmitted from farm to farm, and once infected, all the stock on that farm were destroyed. (After Gibbens & Wilesmith, 2002.)

Figure 12.17 (a) Reported cases of measles in England and Wales from 1948 to 1968, prior to the introduction of mass vaccination. (b) Reported cases of pertussis (whooping cough) in England and Wales from 1948 to 1982. Mass vaccination was introduced in 1956. (After Anderson & May, 1991.)



to the more rapid removal of susceptibles from the population and hence to an earlier end to the epidemic: higher values of R_0 tend to give rise to shorter, sharper epidemic curves. Also, whether the infection disappears altogether (i.e. the epidemic simply ends) depends very largely on the rate at which new susceptibles either move into or are born into the population, since this determines how long the population remains below S_T . If this rate is too low, then the epidemic will indeed simply end. But a sufficiently rapid input of new susceptibles should prolong the epidemic, or even allow the infection to establish endemically in the population after the initial epidemic has passed.

12.6.4 Directly transmitted microparasites: cycles of infection

dynamic patterns of different types of parasite

This leads us naturally to consider the longer term patterns in the dynamics of different types of endemic infection. As described above, the immunity induced by many bacterial and viral infections reduces S , which reduces R_0 , which therefore tends to lead to a decline in the incidence of the infection itself. However, in due course, and before the infection disappears altogether from the population, there is likely to be an influx of new susceptibles into the population, a subsequent increase in S and R_0 , and so on. There is thus a marked tendency with such infections to generate a sequence from ‘many susceptibles (R_0 high)’, to ‘high incidence’, to ‘few susceptibles (R_0 low)’, to ‘low incidence’, to ‘many susceptibles’, etc. – just like any other predator–prey cycle. This undoubtedly underlies the observed cyclic incidence of many human diseases, with the differing lengths of cycle reflecting the differing characteristics of the diseases: measles with peaks every 1 or 2 years (Figure 12.17a), pertussis (whooping cough) every 3–4 years (Figure 12.17b), diphtheria every 4–6 years, and so on (Anderson & May, 1991).

By contrast, infections that do not induce an effective immune response tend to be longer lasting within individual hosts, but also tend not to give rise to the same sort of fluctuations in S and R_0 . Thus, for example, protozoan infections tend to be much less variable (less cyclic) in their prevalence.

12.6.5 Directly transmitted microparasites: immunization programs

Recognizing the importance of critical population sizes also throws light on immunization programs, in which susceptible hosts are rendered nonsusceptible without ever becoming diseased (showing clinical symptoms), usually through exposure to a killed or attenuated pathogen. The direct effects here are obvious: the immunized individual is protected. But, by reducing the number of susceptibles, such programs also have the indirect effect of reducing R_0 . Indeed, seen in these terms, the fundamental aim of an immunization program is clear – to hold the number of susceptibles below S_T so that R_0 remains less than 1. To do so is said to provide ‘herd immunity’.

In fact, a simple manipulation of Equation 12.5 gives rise to a formula for the critical proportion of the population, p_c , that needs to be immunized in order to provide herd immunity (reducing R_0 to a maximum of 1, at most). If we define S_0 as the typical number of susceptibles prior to any immunization and note that S_T is the number still susceptible (not immunized) once the program to achieve $R_0 = 1$ has become fully established, then the proportion immunized is:

$$p_c = 1 - (S_T/S_0). \quad (12.7)$$

The formula for S_T is given in Equation 12.6, whilst that for S_0 , from Equation 12.5, is simply $R_0/\beta L$, where R_0 is the basic reproductive rate of the infection prior to immunization. Hence:

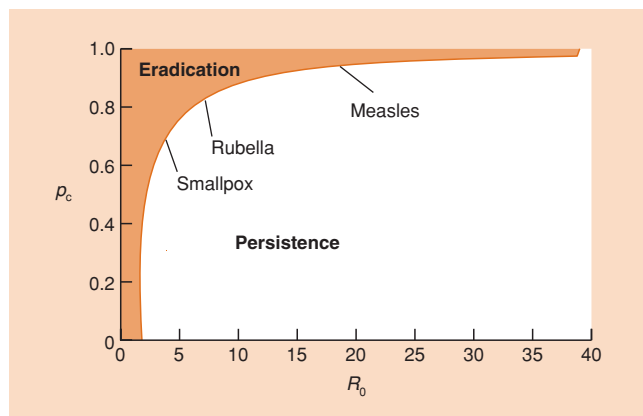


Figure 12.18 The dependence of the critical level of vaccination coverage required to halt transmission, p_c , on the basic reproductive rate, R_0 , with values for some common human diseases indicated. (After Anderson & May, 1991.)

$$p_c = 1 - (1/R_0). \quad (12.8)$$

This reiterates the point that in order to eradicate a disease, it is not necessary to immunize the whole population – just a proportion sufficient to bring R_0 below 1. It also shows that this proportion will be higher the greater the ‘natural’ basic reproductive rate of the disease (without immunization). This general dependence of p_c on R_0 is illustrated in Figure 12.18, with the estimated values for a number of human diseases indicated on it. Note that smallpox, the only disease where in practice immunization seems to have led to eradication, has unusually low values of R_0 and p_c .

12.6.6 Directly transmitted microparasites: frequency-dependent transmission

Suppose, however, that transmission is frequency dependent (see Section 12.4.3), as it is likely to be, for example, with sexually transmitted diseases, where transmission occurs after an infected individual ‘seeks out’ (or is sought out by) a susceptible individual. Then there is no longer the same dependence on the number of susceptibles, and the basic reproductive rate is simply given by:

$$R_0 = \beta' L. \quad (12.9)$$

Here, there is apparently no threshold population size and such infections can therefore persist even in extremely small populations (where, to a first approximation, the chances of sexual contact for an infected host are the same as in large populations).

12.6.7 Crop pathogens: macroparasites viewed as microparasites

Most of plant pathology has been concerned with the dynamics of diseases within crops, and hence with the spread of a disease *within* a generation. Moreover, although most commonly studied plant pathogens are macroparasites in the sense we have defined them, they are typically treated like microparasites in that disease is monitored on the basis of some measure of disease severity – often, the proportion of the population infected (i.e. prevalence). We refer to y_t as the proportion affected by lesions at time t , and hence $(1 - y_t)$ is the proportion of the population without lesions and thus susceptible to infection. It is also usually necessary with plant pathogens to take explicit account of the latent period, length p , between the time when a lesion is initiated and the time when it becomes infectious (spore-forming) itself, in which state it remains for a further period l . Hence, the proportion of the population affected by *infectious* lesions at time t is $(y_{t-p} - y_{t-p-l})$. The rate of increase in the proportion of a plant population affected by lesions (Vanderplank, 1963; Zadoks & Schein, 1979; Gilligan, 1990) may thus be given by:

$$dy_t/dt = D(1 - y_t)(y_{t-p} - y_{t-p-l}), \quad (12.10)$$

which is essentially a βSI formulation, with D the plant pathologists’ version of a transmission coefficient. This gives rise to S-shaped curves for the progress of a disease within a crop that broadly match the data derived from many crop–pathogen systems (Figure 12.19).

In the progress of such infections, plant pathologists recognize three phases.

- 1 The ‘exponential’ phase, when, although the disease is rarely detectable, rapid acceleration of parasite prevalence occurs. This is therefore the phase in which chemical control would be most effective, but in practice it is usually applied in phase 2. The exponential phase is usually considered arbitrarily to end at $y = 0.05$; about the level of infection at which a nonspecialist might detect that an epidemic was developing (the perception threshold).
- 2 The second phase, which extends to $y = 0.5$. (This is sometimes confusingly called the ‘logistic’ phase, although the whole curve is logistic.)
- 3 The terminal phase, which continues until y approaches 1.0. In this phase chemical treatment is virtually useless – yet it is at this stage that the greatest damage is done to the yield of a crop.

On the other hand, some crop diseases are not simply transmitted by the passive spread of infective particles from one host to another. For example, the anther smut fungus, *Ustilago violacea*, is spread between host plants of white campion, *Silene alba*, by

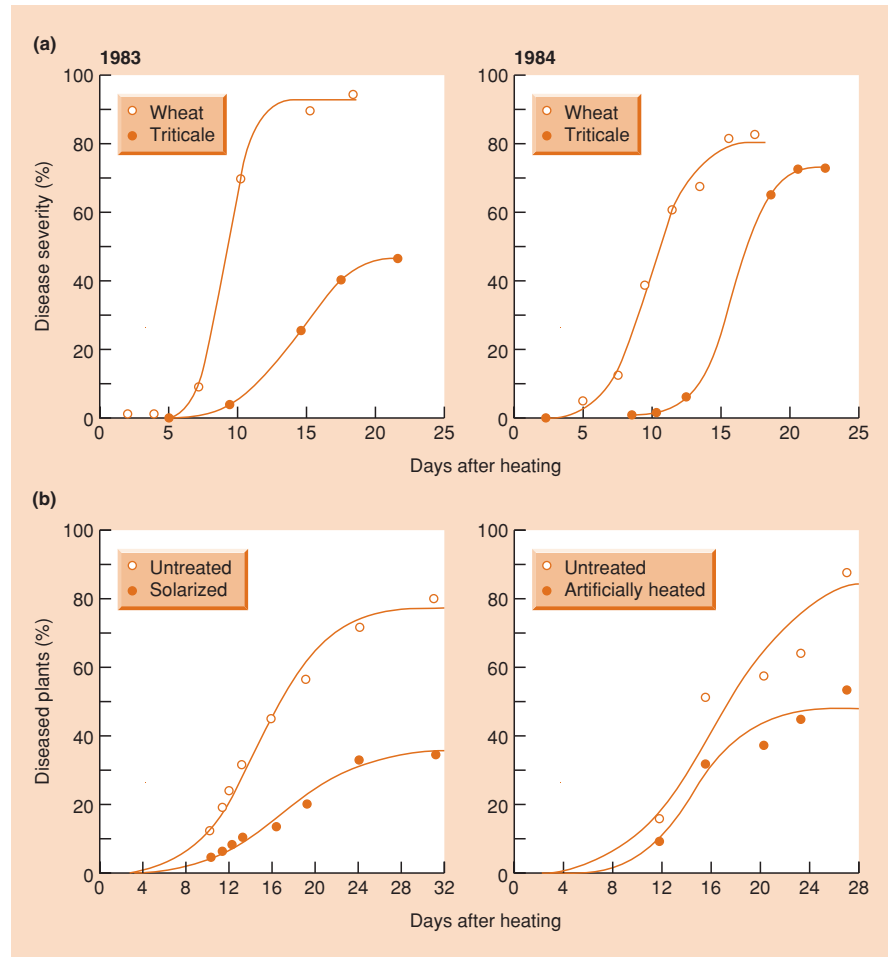


Figure 12.19 ‘S-shaped’ curves of the progress of diseases through crops from an initial inoculum to an asymptotic proportion of the total population infected. (a) *Puccinia recondita* attacking wheat (cultivar Morocco) and triticale (a crop derived from the hybridization of wheat and rye) in 1983 and 1984. (b) *Fusarium oxysporum* attacking tomatoes in experiments comparing untreated and sterilized soil and untreated and artificially heated soil. (After Gilligan, 1990, in which the original data sources and methods of curve-fitting may be found.)

pollinating insects that adjust their flight distances to compensate for changes in plant density, such that the rate of transmission is effectively independent of host density (Figure 12.20a). However, this rate decreases significantly with the proportion of the population that is susceptible: transmission is frequency dependent (Figure 12.20b), favoring, as we have seen, persistence of the disease even in low-density populations. Of course, this is really just another case of frequency-dependent transmission in a sexually transmitted disease – except that sexual contact here is indirect rather than intimate.

12.6.8 Other classes of parasite

vector-borne infections

For microparasites that are spread from one host to another by a vector more generally (where the vector does not compensate for changes in host density as in the above example), the life cycle characteristics of both the host and vector enter into the calculation of R_0 . In particular, the transmission threshold

($R_0 = 1$) is dependent on a ratio of vector : host numbers. For a disease to establish itself and spread, that ratio must exceed a critical level – hence, disease control measures are usually aimed directly at reducing the numbers of vectors, and are aimed only indirectly at the parasite. Many virus diseases of crops, and vector-transmitted diseases of humans and their livestock (malaria, onchocerciasis, etc.), are controlled by insecticides rather than chemicals directed at the parasite; and the control of all such diseases is of course crucially dependent on a thorough understanding of the vector’s ecology.

directly transmitted macroparasites

The effective reproductive rate of a directly transmitted macroparasite (no intermediate host) is directly related to the length of its reproductive period within the host (i.e. again, to L) and to its rate of reproduction (rate of production of infective stages). Both of these are subject to density-dependent constraints that can arise either because of competition between the parasites, or commonly because of the host’s immune response (see Section 12.3.8). Their intensity varies with the distribution of the parasite population between its hosts and, as we have seen,

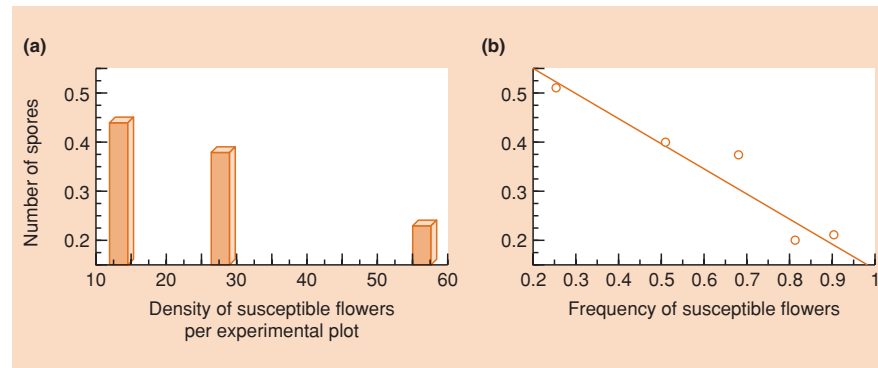


Figure 12.20 Frequency-dependent transmission of a sexually transmitted disease. The number of spores of *Ustilago violacea* deposited per flower of *Silene alba* ($\log_{10}(x + 1)$ transformed) where spores are transferred by pollinating insects. (a) The number is independent of the density of susceptible (healthy) flowers in experimental plots ($P > 0.05$) (and shows signs of decreasing rather than increasing with density, perhaps as the number of pollinators becomes limiting). (b) However, the number decreases with the frequency of susceptibles ($P = 0.015$). (After Antonovics & Alexander, 1992.)

aggregation of the parasites is the most common condition. This means that a very large proportion of the parasites exist at high densities where the constraints are most intense, and this tightly controlled density dependence undoubtedly goes a long way towards explaining the observed stability in prevalence of many helminth infections (such as hookworms and roundworms) even in the face of perturbations induced by climatic change or human intervention (Anderson, 1982).

Most directly transmitted helminths have an enormous reproductive capability. For instance, the female of the human hookworm *Necator* produces roughly 15,000 eggs per worm per day, whilst the roundworm *Ascaris* can produce in excess of 200,000 eggs per worm per day. The critical threshold densities for these parasites are therefore very low, and they occur and persist endemically in low-density human populations, such as hunter-gatherer communities.

indirectly transmitted macroparasites

Density dependence within hosts also plays a crucial role in the epidemiology of indirectly transmitted macroparasites, such as schistosomes. In this case, however, the regulatory constraints can occur in either or both of the hosts: adult worm survival and egg production are influenced in a density-dependent manner in the human host; but also, production of infective stages by the snail (intermediate host) is virtually independent of the number of (different) infective stages that penetrate the snail. Thus, levels of schistosome prevalence tend to be stable and resistant to perturbations from outside influences.

The threshold for the spread of infection depends directly on the abundance of both humans and snails (i.e. a product as opposed to the ratio that was appropriate for vector-transmitted microparasites). This is because transmission in both directions is by means of free-living infective stages. Thus, since it is inappropriate to reduce human abundance, schistosomiasis is often

controlled by reducing snail numbers with molluscicides in an attempt to depress R_0 below unity (the transmission threshold). The difficulty with this approach, however, is that the snails have an enormous reproductive capacity, and they rapidly recolonize aquatic habitats once molluscicide treatment ceases. The limitations imposed by low snail numbers, moreover, are offset to an important extent by the long lifespan of the parasite in humans (L is large): the disease can remain endemic despite wide fluctuations in snail abundance.

12.6.9 Parasites in metapopulations: measles

With host-parasite dynamics, as with other areas of ecology, there is increasing recognition that populations cannot be seen as either homogeneous or isolated. Rather, hosts are usually distributed amongst a series of subpopulations, linked by dispersal between them, and which together comprise a 'metapopulation' (see Section 6.9). Thus, since the argument has already been made (see Section 12.4.1) that each host supports a subpopulation and a host population supports a metapopulation of parasites, host-parasite systems are typically metapopulations of metapopulations.

Such a perspective immediately changes our view of what is required of a host population if it is to support a persistent population of parasites. This is apparent from an analysis of the dynamics of measles in 60 towns and cities in England and Wales from 1944 to 1994: 60 subpopulations comprising an overall metapopulation (Figure 12.21) (Grenfell & Harwood, 1997). Taken as a whole, the metapopulation displayed regular cycles in the number of measles cases and measles was ever-present (Figure 12.21a), at least before widespread vaccination (c. 1968). But amongst the individual subpopulations, only the very largest were not liable to frequent 'stochastic fade-out' (disappearance of the disease when a few remaining infectious individuals fail to pass

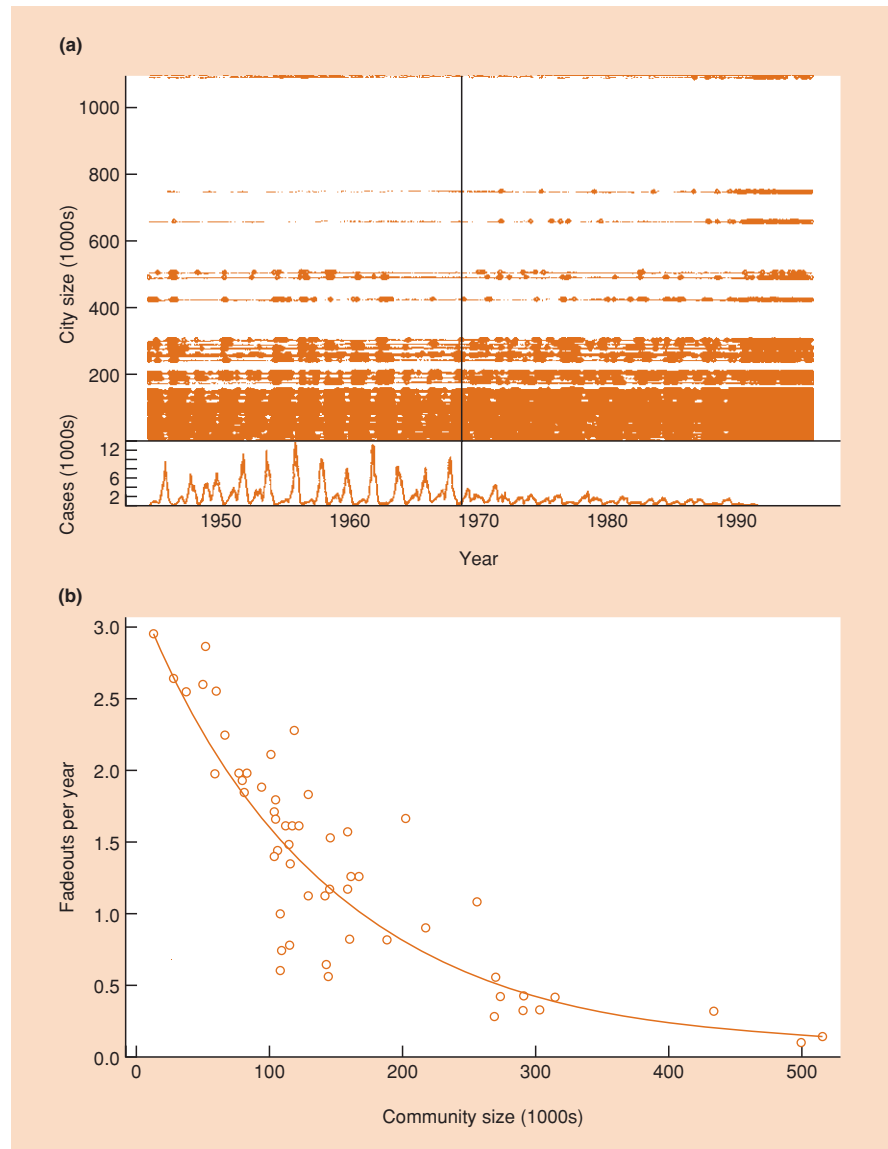


Figure 12.21 (a) The weekly measles notifications for 60 towns and cities in England and Wales, combined, are shown below for the period 1944–94. The vertical line indicates the start of mass vaccination around 1968. The data for the individual towns (town size on the vertical axis) are displayed above as a dot for each week without a measles notification. (b) Persistence of measles in these towns and cities in the prevaccination era (1944–67) as a function of population size. Persistence is measured inversely as the number of ‘fade-outs’ per year, where a fade-out here is defined as a period of three or more weeks without notification, to allow for the underreporting of cases. (After Grenfell & Harwood, 1997.)

it on), especially during the cycle troughs: the idea of a critical population size of around 300,000–500,000 is therefore well supported (Figure 12.21b). Thus, patterns of dynamics may be apparent, and persistence may be predictable, in a metapopulation taken as a whole. But in the individual subpopulations, especially if they are small, the patterns of dynamics and persistence are likely to be far less clear. The measles data set is unusual in that we have information both for the metapopulation and individual subpopulations. In many other cases, it is almost certain that the principle is similar but we have data only for the metapopulation (and do not appreciate the number of fade-outs in smaller parts of it), or we have data only for a subpopulation (and do not appreciate its links to other subpopulations within the larger metapopulation).

12.7 Parasites and the population dynamics of hosts

A key and largely unanswered question in population ecology is what role, if any, do parasites and pathogens play in the dynamics of their hosts? There are data (see Section 12.5) showing that parasites may affect host characteristics of demographic importance (birth and death rates), though even these data are relatively uncommon; and there are mathematical models showing that parasites have the *potential* to have a major impact on the dynamics of their hosts. But the point was also made earlier that it is a big step further to establish that dynamics are actually affected. There are certainly cases where a parasite or pathogen seems, by implication, to reduce the population size of its host. The

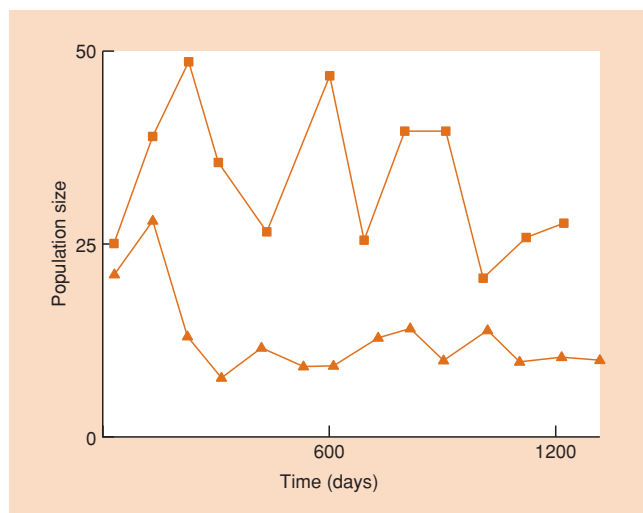


Figure 12.22 Depression of the population size of the flour beetle, *Tribolium castaneum*, infected with the protozoan parasite *Adelina triboli*: (■) uninfected, (▲) infected. (After Park, 1948.)

widespread and intensive use of sprays, injections and medicines in agricultural and veterinary practice all bear witness to the disease-induced loss of yield that would result in their absence. Data sets from controlled, laboratory environments showing reductions in host abundance by parasites have also been available for many years (Figure 12.22). However, good evidence from natural populations is extremely rare. Even when a parasite is present in one population but absent in another, the parasite-free population is certain to live in an environment that is different from that of the infected population; and it is likely also to be infected with some other parasite that is absent from or of low prevalence in the first population. Nevertheless, as we shall see, there are sets of field data in which a parasite is strongly implicated in the detailed dynamics of its host, either as a result of field-scale manipulations, or through using data on the effects of parasites on individual hosts in order to ‘parameterize’ mathematical models that can then be compared with field data.

12.7.1 Coupled (interactive) or modified host dynamics?

First, an important question, even when an effect of a parasite on host dynamics has been demonstrated, is whether the host and parasite interact, such that their dynamics are coupled in the manner usually envisaged for ‘predator–prey’ cycles, or whether the parasite simply modifies the underlying dynamics of the host, without there being any detectable feedback between host and parasite dynamics, and hence without any actual *interaction* between the two. This question has been addressed for the data shown in Figure 12.23 for the stored product moth *Plodia*

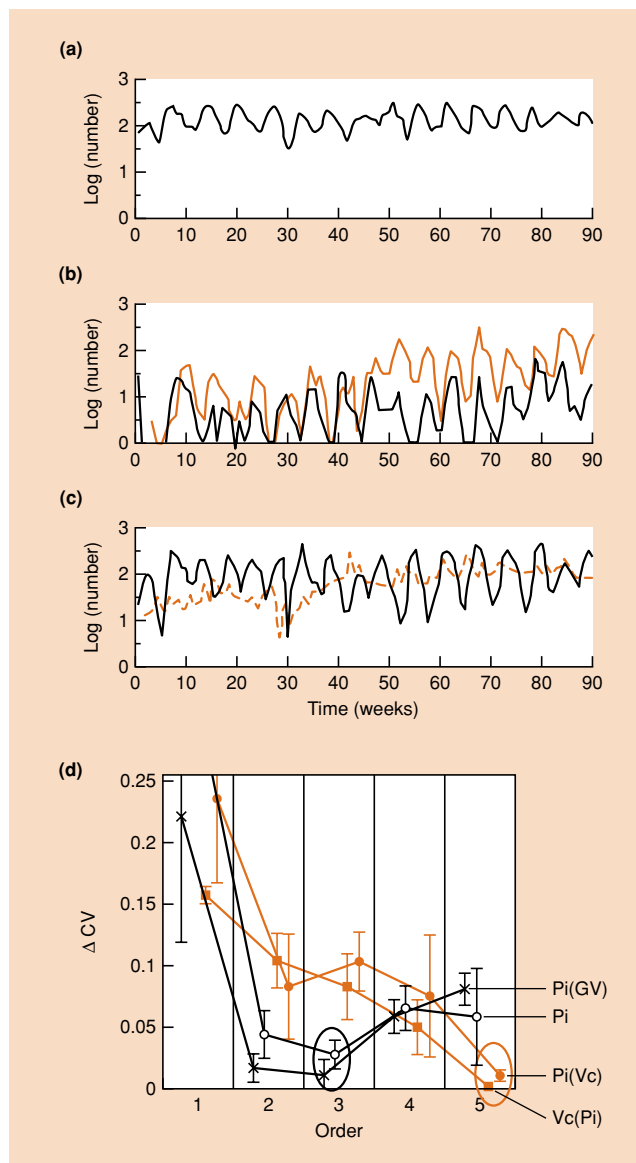


Figure 12.23 Dynamics of the host moth, *Plodia interpunctella* (—) alone (a), in the presence of the parasitoid *Venturia canescens* (—) (b), and in the presence of the *Plodia interpunctella* granulovirus (---) (c). The series show representative replicates of each treatment (out of three) for the first 90 weeks of the experiment. (d) Estimating the dimensionality or ‘order’ of the density dependence of the dynamics for each treatment (all replicates), which is predicted to increase with the number of interacting elements in the system. The lower the value of ΔCV , the better the ‘fit’; error bars represent 1 SE. The best-fitting orders (circled) are three for the host alone (P_i) and the host in the presence of the virus ($P_i(GV)$), but five for the host in the presence of the parasitoid ($P_i(Vc)$), and five too for the parasitoid to which it is coupled ($Vc(P_i)$). (After Bjørnstad *et al.*, 2001.)

interpunctella and its granulovirus (PiGV) (touched on briefly in Section 10.2.5). The dynamics of the host in the presence and absence of the virus are different but only subtly so (Figure 12.23a, c), and detailed statistical analysis is required to try to understand the difference. Put simply, if host dynamics in infected populations are driven by an interaction between *Plodia* and PiGV, then the ‘dimensionality’ of those dynamics (essentially, the complexity of the statistical model required to describe them) should be greater than those of uninfected populations. In fact, although host fecundity was reduced and host development was slowed by the virus, and host abundance was more variable, the dimensionality of the dynamics was unaltered (Figure 12.23d): the virus modulated the vital rates of the host but did not interact with the host nor alter the underlying nature of its dynamics (Bjørnstad *et al.*, 2001). By contrast, when *Plodia* interacted with another natural enemy, the parasitoid *Venturia canescens*, the underlying pattern of ‘generation cycles’ (see Section 10.2.4) remained intact, but this time the dimensionality of the host dynamics was significantly increased (from dimension 3 to 5): the host and parasite interacted.

12.7.2 Red grouse and nematodes

Next we look at the red grouse – of interest both because it is a ‘game’ bird, and hence the focus of an industry in which British landowners charge for the right to shoot at it, and also because it is another species that often, although not always, exhibits regular cycles of abundance (Figure 12.24a). The underlying cause of these cycles has been disputed (Hudson *et al.*, 1998; Lambin *et al.*, 1999; Mougeot *et al.*, 2003), but one mechanism receiving strong support has been the influence of the parasitic nematode, *Trichostrongylus tenuis*, occupying the birds’ gut ceca and reducing survival and breeding production (Figure 12.24b, c).

A model for this type of host–macroparasite interaction is described in Figure 12.25. Its analysis suggests that regular cycles both of host abundance and of mean number of parasites per host will be generated if:

$$\delta > \alpha k. \quad (12.11)$$

Here, δ is the parasite-induced reduction in host fecundity (relatively delayed density dependence: destabilizing), α is the parasite-induced host death rate (relatively direct density dependence: stabilizing), and k is the ‘aggregation parameter’ for the (assumed) negative binomial distribution of parasites amongst hosts. Cycles arise when the destabilizing effects of reduced fecundity outweigh the stabilizing effects of both increased mortality and the aggregation of parasites (providing a ‘partial refuge’ for the hosts) (see Chapter 10). Data from a cyclic study population in the north of England indicate that this condition is indeed satisfied. On the other hand, grouse populations that fail to show regular cycles or show them only very sporadically are often those in which the

nematode cannot properly establish (S_T exceeds typical host abundance) (Dobson & Hudson, 1992; Hudson *et al.*, 1992b).

Such results from models are supportive of a role for the parasites in grouse cycles, but they fall short of the type of ‘proof’ that can come from a controlled experiment. A simple modification of the model in Figure 12.25, however, predicted that if a sufficient proportion (20%) of the population were treated for their nematodes with an anthelmintic, then the cycles would die out. This set the scene for a field-scale experimental manipulation designed to test the parasite’s role (Hudson *et al.*, 1998). In two populations, the grouse were treated with anthelmintics in the expected years of two successive population crashes; in two others, the grouse were treated only in the expected year of one crash; while two further populations were monitored as unmanipulated controls. Grouse abundance was measured as ‘bag records’: the number of grouse shot. It is clear that the anthelmintic had an effect in the experiment (Figure 12.24d), and it is therefore equally clear that the parasites normally have an effect: that is, the parasites affected host dynamics. The precise nature of that effect, however, remains a matter of some controversy. Hudson and his colleagues themselves believed that the experiment demonstrated that the parasites were ‘necessary and sufficient’ for host cycles. Others felt that rather less had been fully demonstrated, suggesting for example that the cycles may have been reduced in amplitude rather than eliminated, especially as the very low numbers normally ‘observed’ in a trough (1 on their logarithmic scale equates to zero) are a result of there being no shooting when abundance is low (Lambin *et al.*, 1999; Tompkins & Begon, 1999). On the other hand, such controversy should not be seen as detracting from the importance of field-scale experiments in investigating the roles of parasites in the dynamics of host populations – nor, indeed, the roles of other factors. For example, a subsequent field manipulation supported the alternative hypothesis that red grouse cycles are the result of density-dependent changes in aggressiveness and the spacing behavior of males (Mougeot *et al.*, 2003). This system is examined again in a general discussion of cycles in Section 14.6.2.

12.7.3 Svalbard reindeer and nematodes

Next, we stay with nematodes but switch to a mammal, the Svalbard reindeer, *Rangifer tarandus plathyrinchus*, on the island of Svalbard (Spitzbergen), north of Norway (Albon *et al.*, 2002). The system is attractive for its simplicity (the effects may be visible, uncluttered by complicating factors): (i) there are no mammalian herbivores competing with the reindeer for food; (ii) there are no mammalian predators; and (iii) the parasite community of the reindeer is itself very simple, dominated by two gastrointestinal nematodes, neither with an alternative host and only one of which, *Ostertagia gruehneri*, has a demonstrable pathogenic effect.

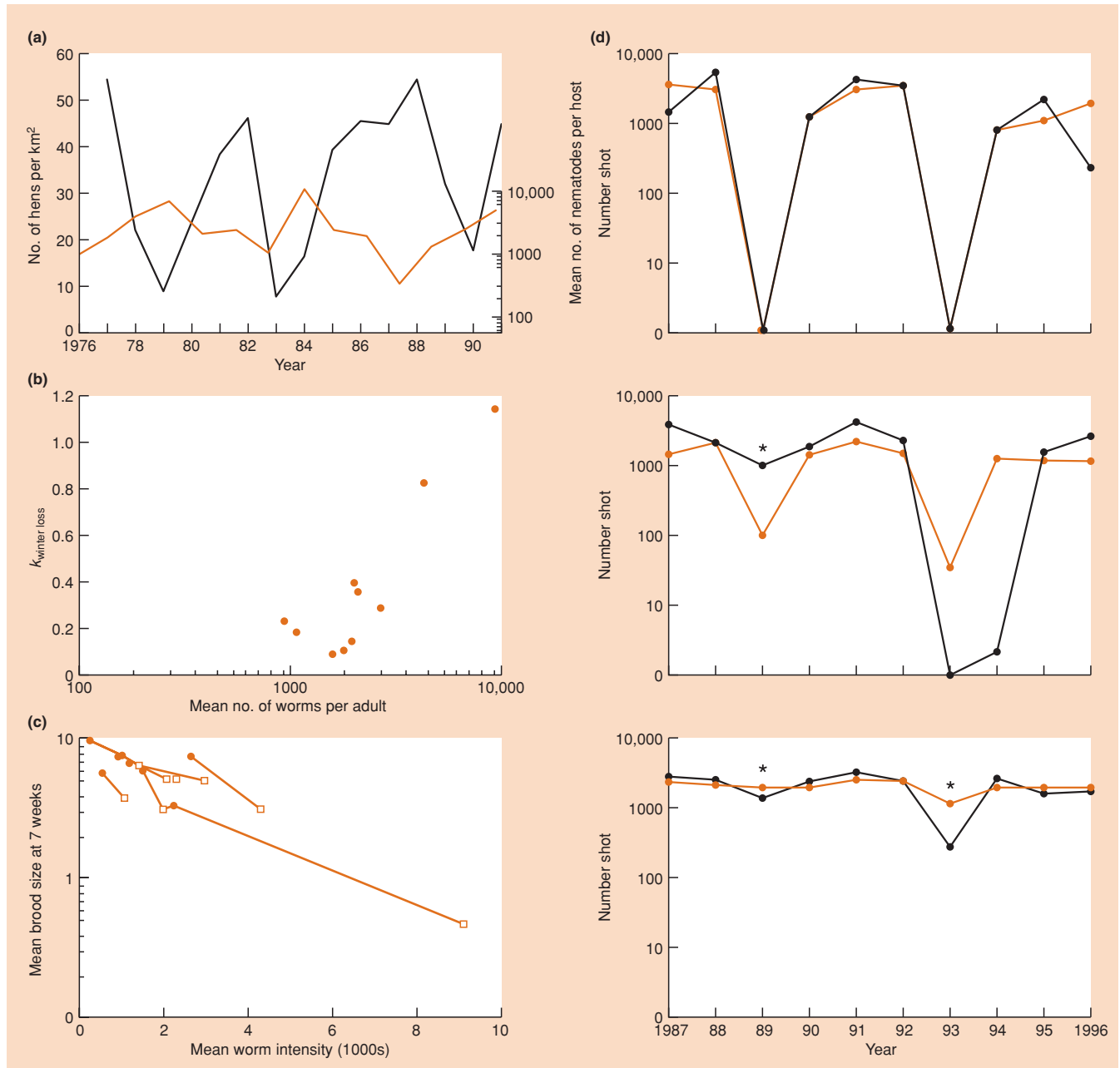
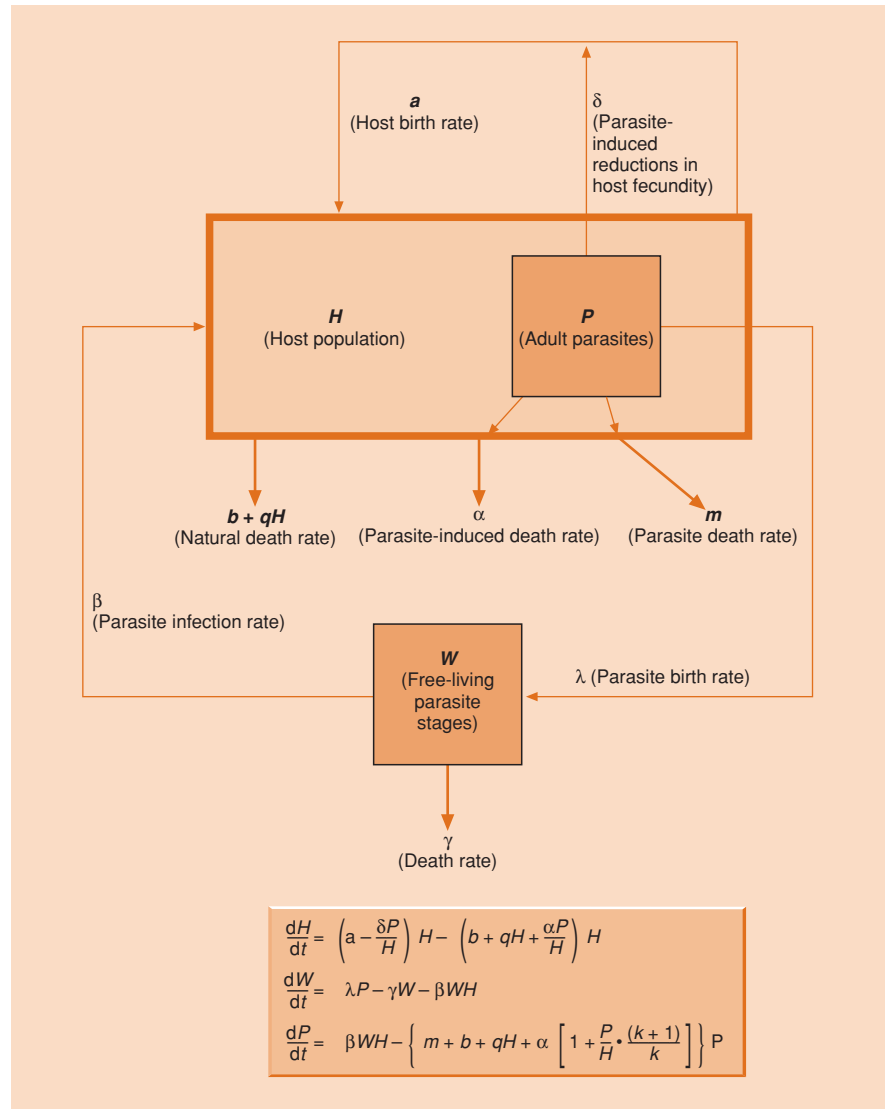


Figure 12.24 (a) Regular cycles in the abundance (breeding hens per km²) of red grouse (—) and the mean number of nematodes, *Trichostrongylus tenuis*, per host (—) at Gunnerside, UK. (b) *Trichostrongylus tenuis* reduces survival in the red grouse: over 10 years (1980–89) winter loss (measured as a k value) increased significantly ($P < 0.05$) with the mean number of worms per adult. (c) *T. tenuis* reduces fecundity in the red grouse: in each of 8 years, females treated with a drug to kill nematodes (●; representing mean values) had fewer worms and larger brood sizes (at 7 weeks) than untreated females (□). ((a–c) after Dobson & Hudson, 1992; Hudson *et al.*, 1992.) (d) Population changes of red grouse, as represented through bag records in two control sites (above), two populations with a single treatment each against nematodes (middle), and two populations with two treatments each (below). Asterisks represent the years of treatment when worm burdens in adult grouse were reduced by an anthelmintic. (After Hudson *et al.*, 1998.)

Figure 12.25 Flow diagram (above) depicting the dynamics of a macroparasitic infection such as the nematode *Trichostrongylus tenuis* in red grouse, where the parasite has free-living infective stages; and (below) the model equations describing those dynamics. Taking the equations in order, they describe: (i) hosts (H) increasing as a result of (density-independent) births (which, however, are reduced at a rate dependent on the average number of parasites per host, P/H), but decreasing as a result of deaths – both natural (density dependent) and induced by the parasite (again dependent on P/H); (ii) free-living parasite stages (W) increasing as a result of being produced by parasites in infected hosts, but decreasing both as a result of death and by being consumed by hosts; and (iii) parasites within hosts (P) increasing as a result of being consumed by hosts, but decreasing as a result of their own death within hosts, of the natural death of the hosts themselves and of disease-induced death of hosts. This final term is dependent on the distribution of parasites amongst hosts – here assumed to follow a negative binomial distribution, parameter k , accounting for the term in square brackets. (After Anderson & May, 1978; Dobson & Hudson, 1992.)



Over a period of 6 years, reindeer were treated with an anthelmintic each spring (April), and the effect of this on pregnancy rates 1 year later, as well as on subsequent calf production, was noted. Infection appeared to have no effect on survival, but untreated (i.e. infected) females had significantly lower pregnancy rates, after year-to-year variation had been accounted for ($X^2_1 = 4.92$, $P = 0.03$; Figure 12.26a), an effect that was maintained in the data on calf production. The extent of this effect increased significantly with increases in the abundance of the nematode in the previous fall ($F_{1,4} = 52.9$, $P = 0.002$; Figure 12.26b). Moreover, the abundance of the nematodes themselves was significantly and positively related to the density of reindeer 2 years earlier (Figure 12.26c). Hence, increases in host abundance appear to lead (after a delay) to increases in parasite abundance; increases in parasite abundance appear to lead (after a further delay) to

reductions in host fecundity; and reductions in host fecundity clearly have the potential to lead to reductions in host abundance.

In order to ask whether this circle was completed in practice, such that the parasite *did* regulate reindeer abundance, these various relationships, along with others, were fed as parameter values into a model of the reindeer–nematode interaction. Results are shown in Figure 12.26d. Three outcomes are possible: either the reindeer population is driven to extinction, or it shows unbounded exponential growth, or it is regulated to the numbers per square kilometer shown in the figure. Encouragingly, within the observed ranges of calf and old reindeer survival, the model predicts reindeer densities very much in line with those observed (around 1–3 km⁻²). In the absence of an effect of the nematode on calf production, the model predicts unbounded growth. Thus, together, field experiments and observations, and a mathematical

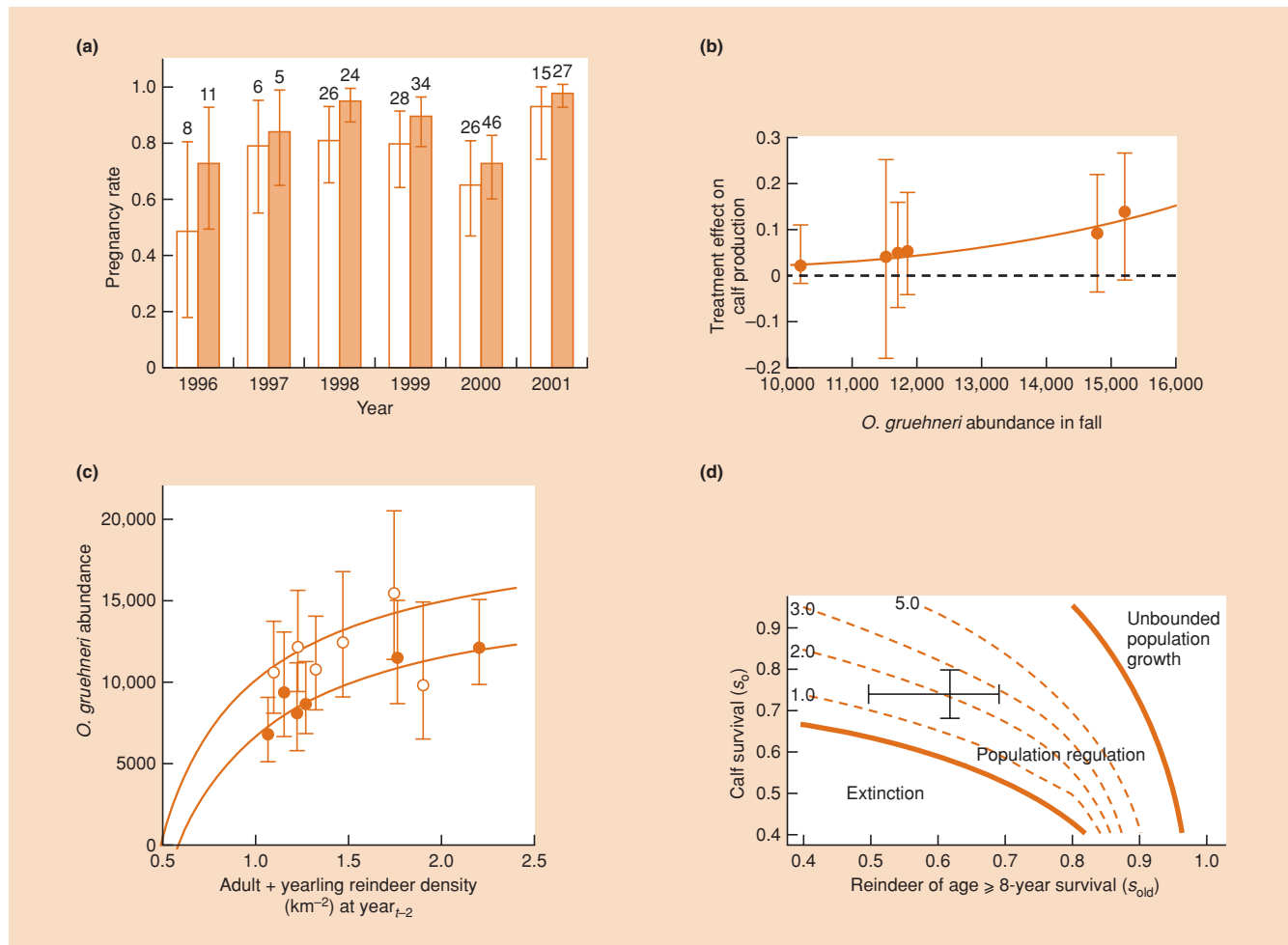


Figure 12.26 (a) The estimated pregnancy rate in April–May in controls (open bars) and reindeer treated with anthelmintics 12 months earlier (shaded bars). Numbers over the bars give the sample size of animals with pregnancy status determined. (b) The difference in the calf production of reindeer treated with anthelmintics in the previous April–May and controls, in relation to the estimated *Ostertagia gruehneri* abundance in October. (c) The estimated *Ostertagia gruehneri* abundance in October in relation (curvilinear regression) to adult and yearling reindeer summer density 2 years earlier at two sites: Colesdalen (●) and Sassendalen (○). Error bars in (a–c) give 95% confidence limits of the estimates. (d) Summary of the output from a model of the Svalbard reindeer population dynamics, using the range of possible values of annual calf survival and the annual survival of reindeer more than or equal to 8 years old. The bold lines give the boundaries between the parameter space where the host population becomes extinct, or is regulated, or shows unbounded growth. The dotted lines give the combination of parameter values in the regulated zone that give an average adult + yearling population density of 1, 2, 3 and 5 reindeer per km². The crossed bars indicate ranges of estimated values. (After Albon *et al.*, 2002.)

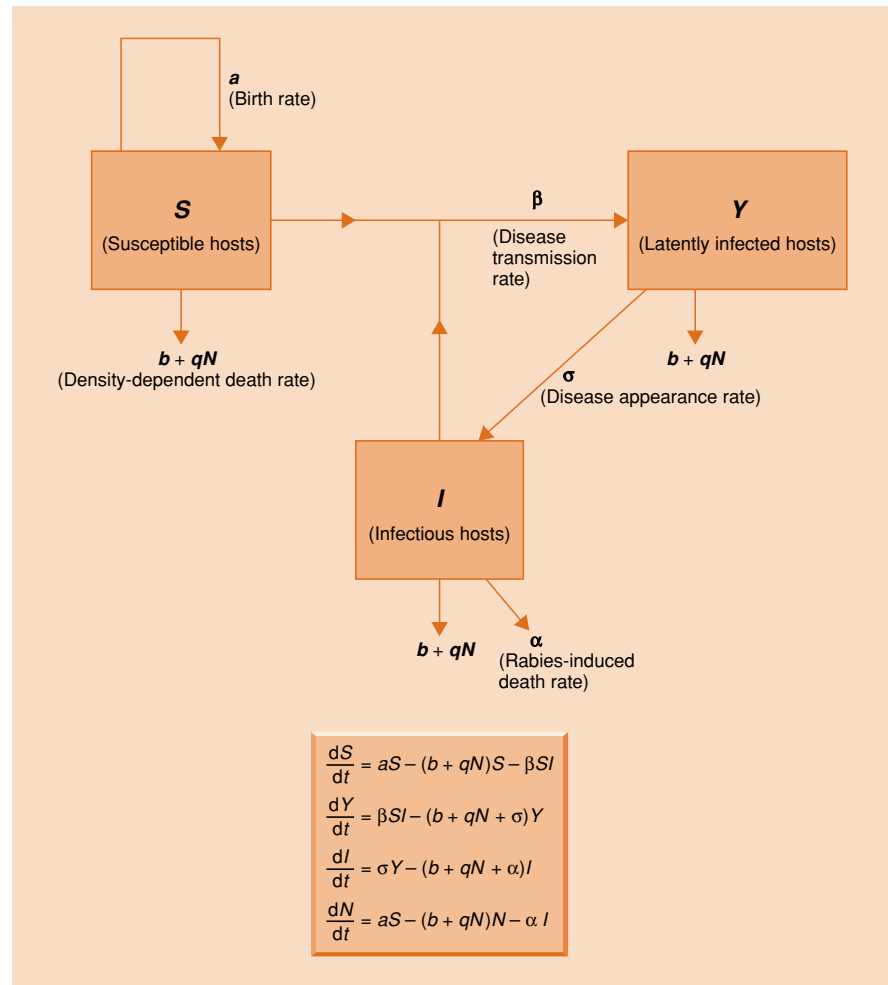
model, provide powerful support for a role of the nematodes in the dynamics of the Svalbard reindeer.

12.7.4 Red foxes and rabies

We turn last to rabies: a directly transmitted viral disease of vertebrates, including humans, that attacks the central nervous system and is much feared both for the unpleasantness of its symptoms and the high probability of death once it has taken hold. In

Europe, recent interest has focused on the interaction between rabies and the red fox (*Vulpes vulpes*). An epidemic in foxes spread westwards and southwards from the Polish–Russian border from the 1940s, and whilst the direct threat to humans is almost certainly slight, there is an economically significant transmission of rabies from foxes to cattle and sheep. The authorities in Great Britain have been especially worried about rabies since the disease has yet to cross the English Channel from mainland Europe, but there has been a strong desire to eliminate rabies from the European mainland too (Pastoret & Brochier, 1999). In this case,

Figure 12.27 Flow diagram (above) depicting the dynamics of a rabies infection of a vertebrate host (such as the fox) and (below) the model equations describing those dynamics. Taking the equations in order, they describe: (i) susceptible hosts (S) increasing as a result of (density-independent) birth from the susceptible class only, but decreasing both as a result of natural (density-dependent) death and also by becoming infected through contact with infectious hosts; (ii) latently infected (noninfectious) hosts (Y) increasing as a result of susceptibles becoming infected, and decreasing both as a result of natural (density-dependent) death and also (as the rabies appears) by becoming converted into infectious hosts; and (iii) infectious hosts (I) increasing as a result of disease development in latently infected hosts, but decreasing as a result of natural and disease-induced mortality. Finally, the equation for the total host population ($N = S + Y + I$) is derived by summing the equations for S , Y and I . (After Anderson *et al.*, 1981.)



we look at the use of a model, first, to capture the observed host–pathogen dynamics in the field (and thus lend credibility to that model) and then to ask whether those dynamics can usefully be manipulated. That is: do we know enough about fox–rabies population dynamics to suggest how further spread of the disease might be prevented and how it might even be eliminated where it already exists?

A simple model of fox–rabies dynamics is described in Figure 12.27. This does indeed seem to capture the essence of the interaction successfully, since, with values for the various biological parameters taken from field data, the model predicts regular cycles of fox abundance and rabies prevalence, around 4 years in length – just like those found in a number of areas where rabies is established (Anderson *et al.*, 1981).

There are two methods that have a realistic chance of controlling rabies in foxes. The first is to kill numbers of them on a continuing basis, so as to hold their abundance below the rabies transmission threshold. The model suggests that this is around 1 km^{-2} , which is itself a helpful piece of information, given credence by the ability of the model to recreate observed

dynamics. As discussed much more fully in Chapter 15 (in the context of harvesting), the problem with repeated culls of this type is that by reducing density they relieve the pressure of intraspecific competition, leading to increases in birth rates and declines in natural death rates. Thus, culling becomes rapidly more problematic the greater the gap between the normal density and the target density (in this case, 1 km^{-2}). Culling may, therefore, be feasible with natural densities of around only 2 km^{-2} . However, since densities in, for instance, Great Britain often average 5 km^{-2} and may reach 50 km^{-2} in some urban areas, culls of a sufficient intensity will usually be unattainable. Culling will typically be of little practical use.

The second potential control method is vaccination – in this case, the placement of oral vaccine in baits to which the foxes are attracted. Such methods can reach around 80% of a fox population. Is that enough? The formula for answering this has already been given as Equation 12.7; the application of which suggests that vaccination should be successful at natural fox densities of up to 5 km^{-2} . Vaccination should therefore be successful, for example, throughout much of Great Britain, but appears to

offer little hope of control in many urban areas. In fact, more than 20 years after the development of the model in Figure 12.27, rabies has still not spread to Great Britain, and the use of ever-improving oral vaccines appears to have halted the spread of rabies in Europe and indeed eliminated it from Belgium, Luxembourg and large parts of France (Pastoret & Brochier, 1999).

12.8 Coevolution of parasites and their hosts

myxomatosis

It may seem straightforward that parasites in a population select for the evolution of more resistant hosts, which in turn select for more infective parasites: a classic coevolutionary arms race. In fact, the process is not necessarily so straightforward, although there are certainly examples where the host and parasite drive one another's evolution. A most dramatic example involves the rabbit and the myxoma virus, which causes myxomatosis. The virus originated in the South American jungle rabbit *Sylvilagus brasiliensis*, where it causes a mild disease that only rarely kills the host. The South American virus, however, is usually fatal when it infects the European rabbit *Oryctolagus cuniculus*. In one of the greatest examples of the biological control of a pest, the myxoma virus was introduced into Australia in the 1950s to control the European rabbit, which had become a pest of grazing lands. The disease spread rapidly in 1950–51, and rabbit populations were greatly reduced – by more than 90% in some places. At the same time, the virus was introduced to England and France, and there too it resulted in huge reductions in the rabbit populations. The evolutionary changes that then occurred in Australia were followed

in detail by Fenner and his associates (Fenner & Ratcliffe, 1965; Fenner, 1983) who had the brilliant research foresight to establish baseline genetic strains of both the rabbits and the virus. They used these to measure subsequent changes in the virulence of the virus and the resistance of the host as they evolved in the field.

When the disease was first introduced to Australia it killed more than 99% of infected rabbits. This 'case mortality' fell to 90% within 1 year and then declined further (Fenner & Ratcliffe, 1965). The virulence of isolates of the virus sampled from the field was graded according to the survival time and the case mortality of control rabbits. The original, highly virulent virus (1950–51) was grade I, which killed 99% of infected laboratory rabbits. Already by 1952 most of the virus isolates from the field were the less virulent grades III and IV. At the same time, the rabbit population in the field was increasing in resistance. When injected with a standard grade III strain of the virus, field samples of rabbits in 1950–51 had a case mortality of nearly 90%, which had declined to less than 30% only 8 years later (Marshall & Douglas, 1961) (Figure 12.28).

This evolution of resistance in the European rabbit is easy to understand: resistant rabbits are obviously favored by natural selection in the presence of the myxoma virus. The case of the virus, however, is subtler. The contrast between the virulence of the myxoma virus in the European rabbit and its lack of virulence in the American host with which it had coevolved, combined with the attenuation of its virulence in Australia and Europe after its introduction, fit a commonly held view that parasites evolve toward becoming benign to their hosts in order to prevent the parasite eliminating its host and thus eliminating its habitat. This view, however, is quite wrong. The parasites favored by natural selection are those with the greatest fitness (broadly, the greatest

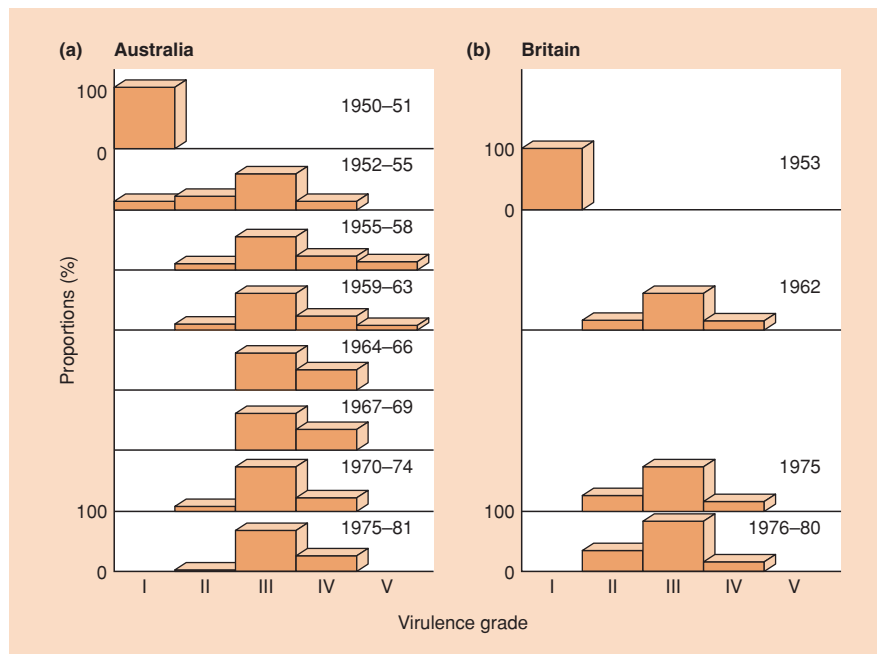


Figure 12.28 (a) The percentages in which various grades of myxoma virus have been found in wild populations of rabbits in Australia at different times from 1951 to 1981. Grade I is the most virulent. (After Fenner, 1983.) (b) Similar data for wild populations of rabbits in Great Britain from 1953 to 1980. (After May & Anderson, 1983; from Fenner, 1983.)

reproductive rate). Sometimes this is achieved through a decline in virulence, but sometimes it is not. In the myxoma virus, an initial decline in virulence was indeed favored – but further declines were not.

The myxoma virus is blood-borne and is transmitted from host to host by blood-feeding insect vectors. In Australia in the first 20 years after its introduction, the main vectors were mosquitoes (especially *Anopheles annulipes*), which feed only on live hosts. The problem for grade I and II viruses is that they kill the host so quickly that there is only a very short time in which the mosquito can transmit them. Effective transmission may be possible at very high host densities, but as soon as densities decline, it is not. Hence, there was selection against grades I and II and in favor of less virulent grades, giving rise to longer periods of host infectiousness. At the other end of the virulence scale, however, the mosquitoes are unlikely to transmit grade V of the virus because it produces very few infective particles in the host skin that could contaminate the vectors' mouthparts. The situation was complicated in the late 1960s when an alternative vector of the disease, the rabbit flea *Spilopsyllus cuniculi* (the main vector in England), was introduced to Australia. There is some evidence that more virulent strains of the virus may be favored when the flea is the main vector (see discussion in Dwyer *et al.*, 1990).

Overall, then, there has been selection in the rabbit–myxomatosis system not for decreased virulence as such, but for *increased transmissibility* (and hence increased fitness) – which happens in this system to be maximized at intermediate grades of virulence. Many parasites of insects rely on killing their host for effective transmission. In these, very high virulence is favored. In yet other cases, natural selection acting on parasites has clearly favored very low virulence: for example, the human herpes simplex virus may do very little tangible harm to its host but effectively gives it lifelong infectiousness. These differences reflect differences in the underlying host–parasite ecologies, but what the examples have in common is that there has been evolution toward increased parasite fitness.

In other cases, coevolution is more definitely antagonistic: increased resistance in the host and increased infectivity in the parasite. A classic example is the interaction between agricultural plants and their pathogens (Burdon, 1987), although in this case the resistant hosts are often introduced by human intervention. There may even be gene-for-gene matching, with a particular virulence allele in the pathogen selecting for a resistant allele in the host, which in turn selects for alleles other than the original allele in the pathogen, and so on. This, moreover, may give rise to polymorphism in the parasite and host, either as a result of different alleles being favored in different subpopulations, or because several alleles are simultaneously in a state of flux within their population, each being favored when they (and their matching allele in the other partner) is rare. In fact, such detailed processes have proved difficult to observe, but this has

bacteria and bacteriophages

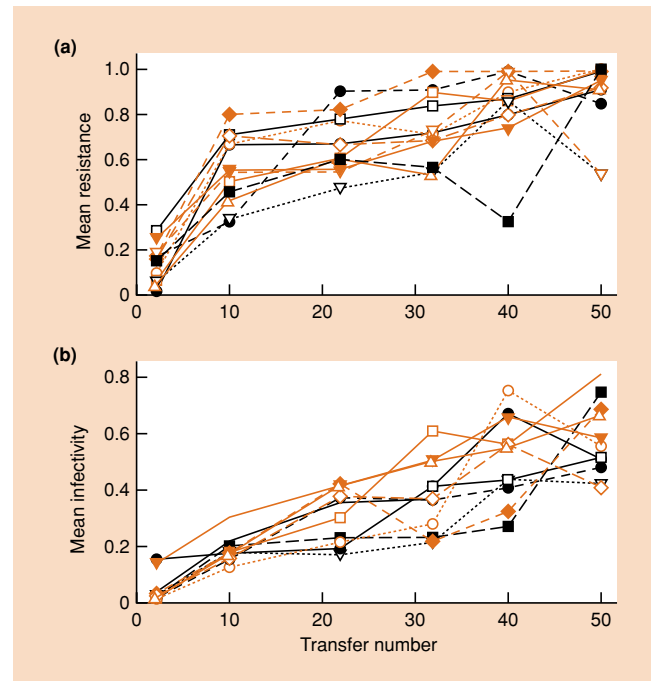


Figure 12.29 (a) Over evolutionary time (1 ‘transfer’ \approx 8 bacterial generations) bacterial resistance to phage increased in each of 12 bacterial replicates. ‘Mean’ resistance was the mean calculated over the 12 phage isolates from the respective time points. (b) Similarly, phage infectivity increased, where ‘mean’ infectivity was calculated over the 12 bacterial replicates. (After Buckling & Rainey, 2002.)

been done with a system comprising the bacterium *Pseudomonas fluorescens* and its viral parasite, the bacteriophage (or phage) SBW25 ϕ 2 (Buckling & Rainey, 2002).

Changes in both the host and parasite were monitored over evolutionary time, as 12 replicate coexisting populations of bacterium and phage were transferred from culture bottle to culture bottle. It is apparent that the bacteria became generally more resistant to the phage at the same time as the phage became generally more infective to the bacteria (Figure 12.29): each was being driven by the directional selection of an arms race. But this was only apparent because any given bacterial strain (from one of the 12 replicates) was tested against all 12 phage strains, and the phage strains were tested similarly. When, at the end of the experiment (Table 12.4), the resistance of each bacterial strain was tested against each phage strain in turn, it was clear that the bacteria were almost always *most* resistant (and often wholly resistant) to the phage strain with which they coevolved. There was therefore extensive evolutionary divergence amongst the strains – or subpopulations – and extensive polymorphism within the metapopulation as a whole.

Thus we close this chapter, appropriately, with another reminder that despite being relatively neglected by ecologists in the past, parasites are increasingly being recognized as major players in both the ecological and the evolutionary dynamics of their hosts.

Table 12.4 For each of 12 bacterial replicates (B1–B12) and their 12 respective phage replicates ($\phi 1$ – $\phi 12$), entries in the table are the proportion of bacteria resistant to the phage at the end of a period of coevolution (50 transfers \approx 400 bacterial generations). Coevolving pairs are shown along the diagonal in bold. Note that bacterial strains are usually most resistant to the phage strain with which they coevolved. (After Buckling & Rainey, 2002.)

Phage replicates	Bacterial replicates											
	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
$\phi 1$	0.8	0.9	1	1	1	1	1	1	0.85	0.85	0.75	0.65
$\phi 2$	0.1	1	0.3	1	0.85	0.25	1	1	0.85	0.9	0.8	0.65
$\phi 3$	0.75	0.75	1	1	1	0.9	1	1	0.85	0.9	0.9	0.65
$\phi 4$	0.15	0.9	0.8	1	0.85	0.6	0.6	1	0.85	1	0.85	0.35
$\phi 5$	0.25	0.9	1	1	1	0.9	1	0.8	0.85	1	0.8	0.65
$\phi 6$	0.2	1	0.85	0.8	0.75	0.8	0.85	0.9	0.85	0.75	0.45	0.25
$\phi 7$	0.2	0.75	0.6	1	0.4	0.45	1	0.9	0.85	1	0.75	0.35
$\phi 8$	0	0.95	0.55	0.95	0.35	0.25	0.8	1	0.85	1	0.7	0.25
$\phi 9$	0	0.7	0.55	0.45	0.7	0.35	1	1	0.85	1	0.5	0.1
$\phi 10$	0	0.7	0.9	0.7	0.55	0.9	1	1	0.7	1	0.5	0.4
$\phi 11$	0	0.5	0.9	0.75	0.7	1	1	0.95	0.75	1	1	0.35
$\phi 12$	0	0.15	0	0.1	0.65	0.35	1	1	0.7	0.8	0.85	0.4

Summary

We begin by defining parasite, infection, pathogen and disease. The diversity of animal and plant parasites is then outlined, based on the distinctions between micro- and macroparasites and between those with direct and those with indirect (vectored) life cycles. The particular case of social and brood parasites (e.g. cuckoos) is also described.

We explain the difference between biotrophic and necrotrophic parasites (pioneer saprotrophs), and we use a discussion of zoonoses (wildlife infections transmissible to man) to illustrate the nature of host specificity amongst parasites.

Hosts are reactive environments: they may resist, or recover, or (in vertebrates) acquire immunity. We describe the contrasting responses in vertebrates to micro- and macroparasites and contrast these in turn with the responses of plants to infection. The costliness of host defense against attack is emphasized. Parasites may also induce profound changes in host growth and behavior.

We explain why it may be difficult to distinguish the effects of intraspecific competition amongst parasites from parasite density-dependent host immune responses, and that patterns associated with interspecific competition are as observable amongst parasites as they are in other organisms.

The distinctions between different types of parasite transmission are outlined and a formal description of transmission dynamics is developed, using the form of the contact rate to distinguish between density- and frequency-dependent transmission, though it is emphasized that these may merely be ends of a spectrum. There may also be great spatial variation in the speed with which infection spreads, either as a result of infectious foci or

because of spatial mixtures of susceptible and resistant species or varieties.

The distribution of parasites within host populations is usually aggregated. This makes it especially important to understand the distinctions between prevalence, intensity and mean intensity.

We discuss the effects of parasites on the survivorship, growth and fecundity of hosts. The effects are often subtle, affecting, for example, interactions of hosts with other species.

We then examine the dynamics of infection within host populations. Key concepts here are the basic reproductive rate, R_0 , the transmission threshold ($R_0 = 1$) and the critical population size. These form a framework for directly transmitted microparasites that sheds light on the kinds of population in which we might expect to find different sorts of infection, on the nature of the epidemic curve of an infection, on the dynamic patterns of different types of parasite, and on the planning of immunization programs based on the principle of ‘herd immunity’.

The dynamics are also outlined of pathogens attacking crops, of vector-borne infections and macroparasites, and of parasites infecting metapopulations of hosts.

We examine the role that parasites and pathogens play in the dynamics of their hosts. We address first the question of whether host and parasite dynamics are coupled, or whether the parasite simply modifies the underlying dynamics of the host, without there being any detectable feedback. A series of case studies then emphasizes that data supporting a role for parasites in the dynamics of their hosts are sparse and often liable to alternative interpretations.

Finally, we consider the coevolution of parasites and their hosts, stressing the absence of any ‘cosy accommodation’, but rather that the selective pressures in both cases – parasite and host – favor maximizing individual fitness.

Chapter 13

Symbiosis and Mutualism



13.1 Introduction: symbionts, mutualists, commensals and engineers

No species lives in isolation, but often the association with another species is especially close: for many organisms, the habitat they occupy is an individual of another species. Parasites live within the body cavities or even the cells of their hosts; nitrogen-fixing bacteria live in nodules on the roots of leguminous plants; and so on. Symbiosis ('living together') is the term that has been coined for such close physical associations between species, in which a 'symbiont' occupies a habitat provided by a 'host'.

In fact, parasites are usually excluded from the category of symbionts, which is reserved instead for interactions where there is at least the suggestion of 'mutualism'. A mutualistic relationship is simply one in which organisms of different species interact to their mutual benefit. It usually involves the direct exchange of goods or services (e.g. food, defense or transport) and typically results in the acquisition of novel capabilities by at least one partner (Herre *et al.*, 1999). Mutualism, therefore, need not involve close physical association: mutualists need not be symbionts. For example, many plants gain dispersal of their seeds by offering a reward to birds or mammals in the form of edible fleshy fruits, and many plants assure effective pollination by offering a resource of nectar in their flowers to visiting insects. These are mutualistic interactions but they are not symbioses.

mutualism: reciprocal exploitation not a cosy partnership

It would be wrong, however, to see mutualistic interactions simply as conflict-free relationships from which nothing but good things flow for both partners. Rather, current evolutionary thinking views mutualisms as cases of reciprocal exploitation where, none the less, each partner is a *net* beneficiary (Herre & West, 1997).

Nor are interactions in which one species provides the habitat for another necessarily either mutualistic (both parties benefit:

'+ +') or parasitic (one gains, one suffers: '+ -'). In the first place, it may simply not be possible to establish, with solid data, that each of the participants either benefits or suffers. In addition, though, there are many 'interactions' between two species in which the first provides a habitat for the second, but there is no real suspicion that the first either benefits or suffers in any measurable way as a consequence. Trees, for example, provide habitats for the many species of birds, bats and climbing and scrambling animals that are absent from treeless environments. Lichens and mosses develop on tree trunks, and climbing plants such as ivy, vines and figs, though they root in the ground, use tree trunks as support to extend their foliage up into a forest canopy. Trees are therefore good examples of what have been called ecological or ecosystem 'engineers' (Jones *et al.*, 1994). By their very presence, they create, modify or maintain habitats for others. In aquatic communities, the solid surfaces of larger organisms are even more important contributors to biodiversity. Seaweeds and kelps normally grow only where they can be anchored on rocks, but their fronds are colonized in turn by filamentous algae, by tube-forming worms (*Spirorbis*) and by modular animals such as hydroids and bryozoans that depend on seaweeds for anchorage and access to resources in the moving waters of the sea.

More generally, many of these are likely to be examples of commensal 'interactions' (one partner gains, the other is neither harmed nor benefits: '+ 0'). Certainly, those cases where the harm to the host of a 'parasite' or the benefit to a 'mutualist' cannot be established should be classified as commensal or 'host-guest', bearing in mind that, like guests under other circumstances, they may be unwelcome when the hosts are ill or distressed! Commensals have received far less study than parasites and mutualists, though many of them have ways of life that are quite as specialized and fascinating.

Mutualisms themselves have often been neglected in the past compared to other types of interaction, yet mutualists compose most of the world's biomass. Almost all the plants that dominate

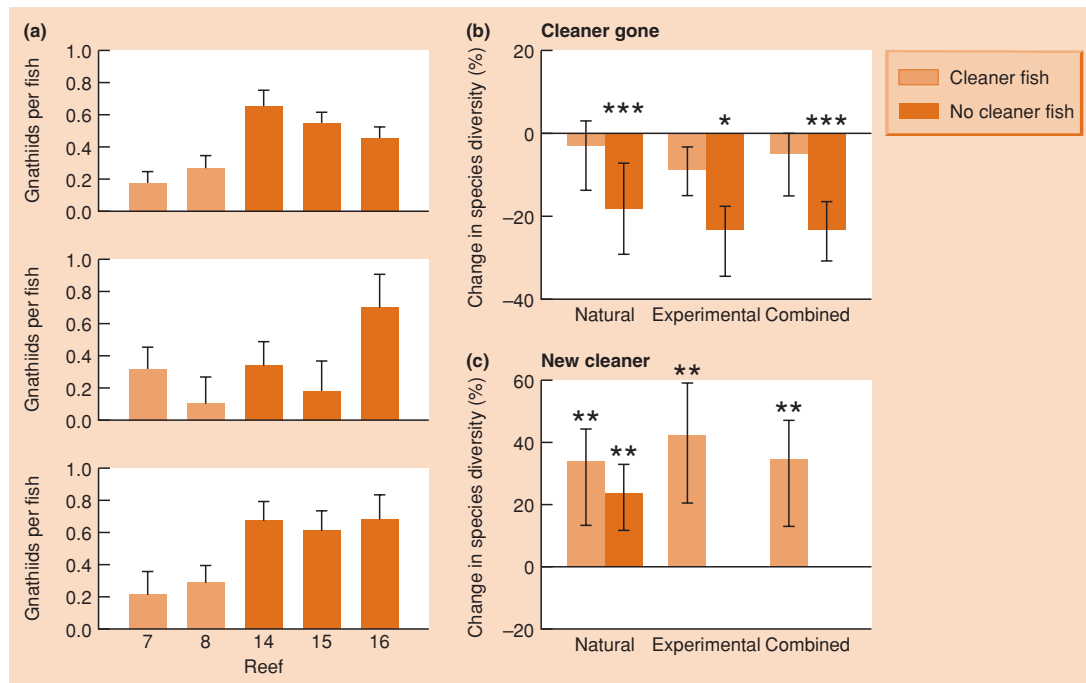


Figure 13.1 (a) Cleaner fish really do clean their clients. The mean number of gnathiid parasites per client (*Hemigymnus melapterus*) at five reefs, from three of which (14, 15 and 16) the cleaners (*Labroides dimidiatus*) were experimentally removed. In a ‘long-term’ experiment, clients without cleaners had more parasites after 12 days (upper panel: $F = 17.6$, $P = 0.02$). In a ‘short-term’ experiment, clients without cleaners did not have significantly more parasites at dawn after 12 h (middle panel: $F = 1.8$, $P = 0.21$), presumably because cleaners do not feed at night, but the difference was significant after a further 12 h of daylight (lower panel: $F = 11.6$, $P = 0.04$). Bars represent standard errors. (After Grutter, 1999.) (b) Cleaners increase reef fish diversity. The percentage change in the number of fish species present following natural or experimental loss of a cleaner fish, *L. dimidiatus*, from a reef patch (or the two treatments combined), in the short term (2–4 weeks, light bars) and the long term (4–20 months, dark bars). (c) The percentage change in the number of fish species present following natural or experimental immigration of a cleaner fish, *L. dimidiatus*, into a reef patch (or the two treatments combined), in the short term (2–4 weeks, light bars) and the long term (4–20 months, dark bars). The columns and error bars represent medians and interquartiles. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (After Bshary, 2003.)

grasslands, heaths and forests have roots that have an intimate mutualistic association with fungi. Most corals depend on the unicellular algae within their cells, many flowering plants need their insect pollinators, and many animals carry communities of microorganisms within their guts that they require for effective digestion.

The rest of this chapter is organised as a progression. We start with mutualisms in which no intimate symbiosis is involved. Rather, the association is largely behavioral: that is, each partner behaves in a manner that confers a net benefit on the other. By Section 13.5, when we discuss mutualisms between animals and the microbiota living in their guts, we will have moved on to closer associations (one partner living within the other), and in Sections 13.6–13.10 we examine still more intimate symbioses in which one partner enters between or within another’s cells. In Section 13.11 we interrupt the progression to look briefly at mathematical

models of mutualisms. Then, finally, in Section 13.12 – for completeness, though the subject is not strictly ‘ecological’ – we examine the idea that various organelles have entered into such intimate symbioses within the cells of their many hosts that it has ceased to be sensible to regard them as distinct organisms.

13.2 Mutualistic protectors

13.2.1 Cleaner and client fish

‘Cleaner’ fish, of which at least 45 species have been recognized, feed on ectoparasites, bacteria and necrotic tissue from the body surface of ‘client’ fish. Indeed, the cleaners often hold territories with ‘cleaning stations’ that their clients visit – and visit more often when they carry many parasites. The cleaners gain a food source

and the clients are protected from infection. In fact, it has not always proved easy to establish that the clients benefit, but in experiments off Lizard Island on Australia's Great Barrier Reef, Grutter (1999) was able to do this for the cleaner fish, *Labroides dimidiatus*, which eats parasitic gnathiid isopods from its client fish, *Hemigymnus melapterus*. Clients had significantly (3.8 times) more parasites 12 days after cleaners were excluded from caged enclosures (Figure 13.1a, top panel); but even in the short term (up to 1 day), although removing cleaners, which only feed during daylight, had no effect when a check was made at dawn (middle panel), this led to there being significantly (4.5 times) more parasites following a further day's feeding (lower panel).

Moreover, further experiments using the same cleaner fish, but at a Red Sea reef in Egypt, emphasized the community-wide importance of these cleaner–client interactions (Bshary, 2003). When cleaners either left a reef patch naturally (so the patch had no cleaner) or were experimentally removed, the local diversity (number of species) of reef fish dropped dramatically, though this was only significant after 4–20 months, not after 2–4 weeks (Figure 13.1b). However, when cleaners either moved into a cleanerless patch naturally or were experimentally added, diversity increased significantly even within a few weeks (Figure 13.1c). Intriguingly, these effects applied not only to client species but to nonclients too.

In fact, several behavioral mutualisms are found amongst the inhabitants of tropical coral reefs (where the corals themselves are mutualists – see Section 13.7.1). The clown fish (*Amphiprion*), for example, lives close to a sea anemone (e.g. *Physobranchia*, *Radianthus*) and retreats amongst the anemone's tentacles when-

ever danger threatens. Whilst within the anemone, the fish gains a covering of mucus that protects it from the anemone's stinging nematocysts (the normal function of the anemone slime is to prevent discharge of nematocysts when neighboring tentacles touch). The fish derives protection from this relationship, but the anemone also benefits because clown fish attack other fish that come near, including species that normally feed on the sea anemones.

13.2.2 Ant–plant mutualisms

The idea that there are mutualistic relationships between plants and ants was put forward by Belt (1874) after observing the behavior of aggressive ants on species of *Acacia* with swollen thorns in Central America. This relationship was later described more fully by Janzen (1967) for the Bull's horn acacia (*Acacia cornigera*) and its associated ant, *Pseudomyrmex ferruginea*. The plant bears hollow thorns that are used by the ants as nesting sites; its leaves have protein-rich 'Beltian bodies' at their tips (Figure 13.2) which the ants collect and use for food; and it has sugar-secreting nectaries on its vegetative parts that also attract the ants. The ants, for their part, protect these small trees from competitors by actively snipping off shoots of other species and also protect the plant from herbivores – even large (vertebrate) herbivores may be deterred.

In fact, ant–plant mutualisms appear to have evolved many times (even repeatedly in the same family of plants); and nectaries are present on

effects at the community level, too

do the plants benefit?

(a)



(b)



Figure 13.2 Structures of the Bull's horn acacia (*Acacia cornigera*) that attract its ant mutualist. (a) Protein-rich Beltian bodies at the tips of the leaflets (© Oxford Scientific Films/Michael Fogden). (b) Hollow thorns used by the ants as nesting sites (© Visuals Unlimited/C. P. Hickman).

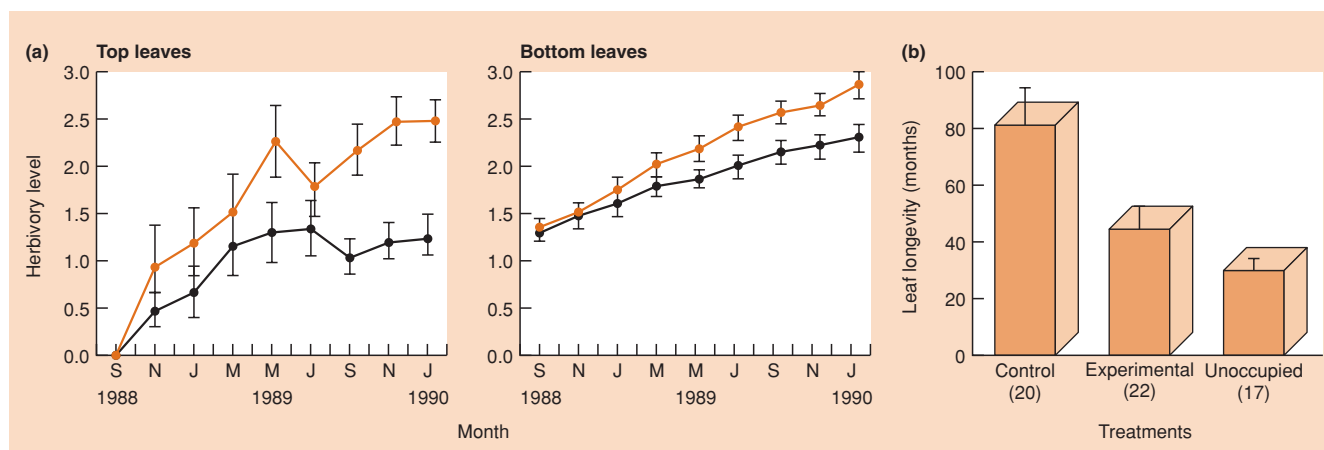


Figure 13.3 (a) The intensity of leaf herbivory on plants of *Tachigali myrmecophila* naturally occupied by the ant *Pseudomyrmex concolor* (●, $n = 22$) and on plants from which the ants had been experimentally removed (●, $n = 23$). Bottom leaves are those present at the start of the experiment and top leaves are those emerging subsequently. (b) The longevity of leaves on plants of *T. myrmecophila* occupied by *P. concolor* (control) and from which the ants were experimentally removed or from which the ants were naturally absent. Error bars \pm standard error. (After Fonseca, 1994.)

the vegetative parts of plants of at least 39 families and in many communities throughout the world. Nectaries on or in flowers are easily interpreted as attractants for pollinators. But the role of extrafloral nectaries on vegetative parts is less easy to establish. They clearly attract ants, sometimes in vast numbers, but carefully designed and controlled experiments are necessary to show that the plants themselves benefit, such as the study of the Amazonian canopy tree *Tachigali myrmecophila*, which harbors the stinging ant *Pseudomyrmex concolor* in specialized hollowed-out structures (Figure 13.3). The ants were removed from selected plants;

these then bore 4.3 times as many phytophagous insects as control plants and suffered much greater herbivory. Leaves on plants that carried a population of ants lived more than twice as long as those on unoccupied plants and nearly 1.8 times as long as those on plants from which ants had been deliberately removed.

Mutualistic relationships, in this case between individual ant and plant species, should not, however, be viewed in isolation – a theme that will recur in this chapter. Palmer *et al.* (2000), for example, studied competition

competition amongst mutualistic ants

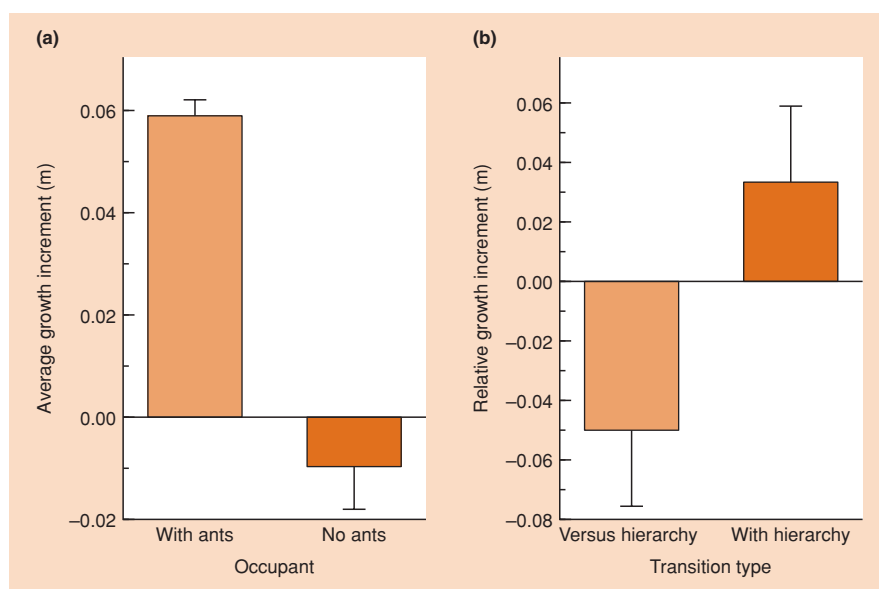


Figure 13.4 (a) Average growth increment was significantly greater ($P < 0.0001$) for *Acacia drepanolobium* trees continually occupied by ants ($n = 651$) than for uninhabited trees ($n = 126$). 'Continually occupied' trees were occupied by ant colonies at both an initial survey and one 6 months later. Uninhabited trees were vacant at the time of both surveys. (b) Relative growth increments were significantly greater ($P < 0.05$) for trees undergoing transitions in ant occupancy in the direction of the ants' competitive hierarchy ($n = 85$) than for those against the hierarchy ($n = 48$). Growth increment was determined relative to trees occupied by the same ant species when these ants were not displaced. Error bars show standard errors. (After Palmer *et al.*, 2000).

amongst four species of ant that have mutualistic relationships with *Acacia drepanolobium* trees in Laikipia, Kenya, nesting within the swollen thorns and feeding from the nectaries at the leaf bases. Experimentally staged conflicts and natural take-overs of plants both indicated a dominance hierarchy among the ant species. *Crematogaster sjostedti* was the most dominant, followed by *C. mimosae*, *C. nigriceps* and *Tetraponera penzigi*. Irrespective of which ant species had colonized a particular acacia tree, occupied trees tended to grow faster than unoccupied trees (Figure 13.4a). This confirmed the mutualistic nature of the interactions overall. But more subtly, changes in ant occupancy in the direction of the dominance hierarchy (take-over by a more dominant species) occurred on plants that grew faster than average, whereas changes in the opposite direction to the hierarchy occurred on plants that grew more slowly than average (Figure 13.4b).

These data therefore suggest that take-overs are rather different on fast and slow growing trees, though the details remain speculative. It may be, for example, that trees that grow fastest also produce ant ‘rewards’ at the greatest rate and are actively chosen by the dominant ant species; whereas slow growing trees are more readily abandoned by dominant species, with their much greater demands for resources. Alternatively, competitively superior ant species may be able to detect and preferentially colonize faster growing trees. What is clear is that these mutualistic interactions are not cosy relationships between pairs of species that we can separate from a more tangled web of interactions. The costs and benefits accruing to the different partners vary in space and time, driving complex dynamics amongst the competing ant species that in turn determine the ultimate balance sheet for the acacias.

Ant–plant interactions are reviewed by Heil and McKey (2003).

13.3 Culture of crops or livestock

13.3.1 Human agriculture

At least in terms of geographic extent, some of the most dramatic mutualisms are those of human agriculture. The numbers of individual plants of wheat, barley, oats, corn and rice, and the areas these crops occupy, vastly exceed what would have been present if they had not been brought into cultivation. The increase in human population since the time of hunter–gatherers is some measure of the reciprocal advantage to *Homo sapiens*. Even without doing the experiment, we can easily imagine the effect the extinction of humans would have on the world population of rice plants or the effect of the extinction of rice plants on the population of humans. Similar comments apply to the domestication of cattle, sheep and other mammals.

Similar ‘farming’ mutualisms have developed in termite and especially ant societies, where the farmers may protect individuals they exploit from competitors and predators and may even move or tend them.

13.3.2 Farming of insects by ants

Ants farm many species of aphids (homopterans) in return for sugar-rich secretions of honeydew. The ‘flocks’

farmed aphids: do they pay a price?

of aphids benefit through suffering lower mortality rates caused by predators, showing increased feeding and excretion rates, and forming larger colonies. But it would be wrong, as ever, to imagine that this is a cosy relationship with nothing but benefits on both sides: the aphids are being manipulated – is there a price that they pay to be entered on the other side of the balance sheet (Stadler & Dixon, 1998)? This question has been addressed for colonies of the aphid *Tuberculatus quercicola* attended by the red wood ant *Formica yessensis* on the island of Hokkaido, northern Japan (Yao *et al.*, 2000). As expected, in the presence of predators, aphid colonies survived significantly longer when attended by ants than when ants were excluded by smearing ant repellent at the base of the oak trees on which the aphids lived (Figure 13.5a). However, there were also costs for the aphids: in an environment from which predators were excluded, and the effects of ant attendance on aphids could thus be viewed in isolation, ant-attended aphids grew less well and were less fecund than those where ants as well as predators were excluded (Figure 13.5b).

Another classic farming mutualism is that between ants and many species of lycaenid butterfly. In a number of cases, young lycaenid caterpillars feed

ants and blue butterflies

on their preferred food plants usually until their third or fourth instar, when they expose themselves to foraging ant workers that pick them up and carry them back to their nests – the ants ‘adopt’ them. There, the ants ‘milk’ a sugary secretion from a specialized gland of the caterpillars, and in return protect them from predators and parasitoids throughout the remainder of their larval and pupal lives. On the other hand, in other lycaenid–ant interactions the evolutionary balance is rather different. The caterpillars produce chemical signals mimicking chemicals produced by the ants, inducing the ants to carry them back to their nests and allowing them to remain there. Within the nests, the caterpillars may either act as social parasites (‘cuckoos’, see Section 12.2.3), being fed by the ants (e.g. the large-blue butterfly *Maculinea rebeli*, which feeds on the crossleaved gentian, *Gentiana cruciata*, and whose caterpillars mimic the larvae of the ant *Myrmica schenkii*), or they may simply prey upon the ants (e.g. another large-blue, *M. arion*, which feeds on wild thyme, *Thymus serpyllum*) (Elmes *et al.*, 2002).

13.3.3 Farming of fungi by beetles and ants

Much plant tissue, including wood, is unavailable as a direct source of food to most animals because they lack the enzymes that can digest cellulose and lignins (see Sections 3.7.2 and 11.3.1). However, many fungi possess these enzymes, and an

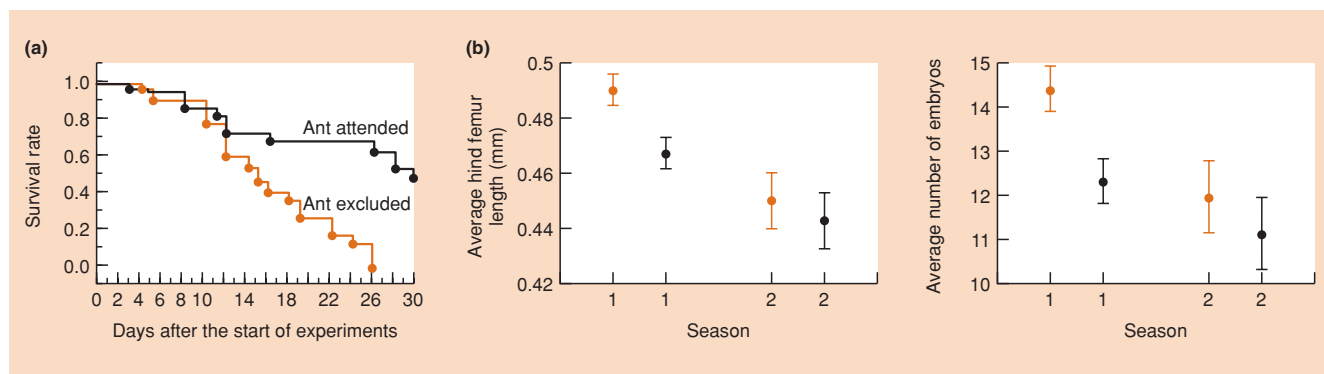


Figure 13.5 (a) Ant-excluded colonies of the aphid *Tuberculatus quercicola* were more likely to become extinct than those attended by ants ($X^2 = 15.9$, $P < 0.0001$). (b) But in the absence of predators, ant-excluded colonies perform better than those attended by ants. Shown are the averages for aphid body size (hind femur length; $F = 6.75$, $P = 0.013$) and numbers of embryos ($F = 7.25$, $P = 0.010$), \pm SE, for two seasons (July 23 to August 11, 1998 and August 12 to August 31, 1998) in a predator-free environment. ●, ant-excluded treatment; ●, ant-attended treatment. (After Yao *et al.*, 2000.)

animal that can eat such fungi gains indirect access to an energy-rich food. Some very specialized mutualisms have developed between animal and fungal decomposers. Beetles in the group *Scolytidae* tunnel deep into the wood of dead and dying trees, and fungi that are specific for particular species of beetle grow in these burrows and are continually grazed by the beetle larvae. These ‘ambrosia’ beetles may carry inocula of the fungus in their digestive tract, and some species bear specialized brushes of hairs on their heads that carry the spores. The fungi serve as food for the beetle and in turn depend on it for dispersal to new tunnels.

Fungus-farming ants are found only in the New World, and the 210 described species appear to have evolved from a common ancestor: that is, the trait has appeared just once in evolution. The more ‘primitive’ species typically use dead vegetative debris as well as insect feces and corpses to manure their gardens; the genera *Trachymyrmex* and *Sericomyrmex* typically use dead vegetable matter; whereas species of the two most derived (evolutionarily ‘advanced’) genera, *Acromyrmex* and *Atta*, are ‘leaf-cutters’ using mostly fresh leaves and flowers (Currie, 2001). Leaf-cutting ants are the most remarkable of the fungus-farming ants. They excavate 2–3-liter cavities in the soil, and in these a basidiomycete fungus is cultured on leaves that are cut from neighboring vegetation (Figure 13.6). The ant colony may depend absolutely on the fungus for the nutrition of their larvae. Workers lick the fungus colonies and remove specialized swollen hyphae, which are aggregated into bite-sized ‘staphylae’. These are fed to the larvae and this ‘pruning’ of the fungus may stimulate further fungal growth. The fungus gains from the association: it is both fed and dispersed by leaf-cutting ants and has never been found outside their nests. The reproductive female ant carries her last meal as a culture when she leaves one colony to found another.

Most phytophagous insects have very narrow diets – indeed, the vast majority of insect herbivores are strict monophages (see Section 9.5). The leaf-cutting ants are remarkable amongst insect herbivores in their polyphagy. Ants from a nest of *Atta cephalotes* harvest from 50 to 77% of the plant species in their neighborhood; and leaf-cutting ants generally may harvest 17% of total leaf production in tropical rainforest and be the ecologically dominant herbivores in the community. It is their polyphagy that gives them this remarkable status. In contrast to the *A. cephalotes* adults though, the larvae appear to be extreme dietary specialists, being restricted to nutritive bodies (gongylidia) produced by the fungus *Attamyces bromatificus*, which the adults cultivate and which decompose the leaf fragments (Cherrett *et al.*, 1989).

Moreover, just as human farmers may be plagued by weeds, so fungus-farming ants have to contend with other species of fungi that may devastate their crop. Fungal pathogens of the genus *Escovopsis* are specialized (never found other than in fungus gardens) and virulent: in one experiment, nine of 16 colonies of the leaf-cutter *Atta colombica* that were treated with heavy doses of *Escovopsis* spores lost their garden within 3 weeks of treatment (Currie, 2001). But the ants have another mutualistic association to help them: a filamentous actinomycete bacterium associated with the surface of the ants is dispersed to new gardens by virgin queens on their nuptial flight, and the ants may even produce chemicals that promote the actinomycete’s growth. For its part, the actinomycete produces an antibiotic with specialized and potent inhibitory effects against *Escovopsis*. It even appears to protect the ants themselves from pathogens and to promote the growth of the farmed fungi (Currie, 2001). *Escovopsis* therefore has ranged

leaf-cutting ants:
remarkably
polyphagous

ants, farmed fungi
and actinomycetes: a
three-way mutualism



Figure 13.6 (a) Partially excavated nest of the leaf-cutting ant *Atta vollenweideri* in the Chaco of Paraguay. The above-ground spoil heap excavated by the ants extended at least 1 m below the bottom of the excavation. (b) Queen of *A. cephalotes* (with an attendant worker on her abdomen) on a young fungus garden in the laboratory, showing the cell-like structure of the garden with its small leaf fragments and binding fungal hyphae. (Courtesy of J. M. Cherrett.)

against it not just two two-species mutualisms but a three-species mutualism amongst ants, farmed fungi and actinomycetes.

13.4 Dispersal of seeds and pollen

13.4.1 Seed dispersal mutualisms

Very many plant species use animals to disperse their seeds and pollen. About 10% of all flowering plants possess seeds or fruits that bear hooks, barbs or glues that become attached to the hairs, bristles or feathers of any animal that comes into contact with them. They are frequently an irritation to the animal,

which often cleans itself and removes them if it can, but usually after carrying them some distance. In these cases the benefit is to the plant (which has invested resources in attachment mechanisms) and there is no reward to the animal.

Quite different are the true mutualisms between higher plants and the **fruits** and other animals that feed on the fleshy fruits and disperse the seeds. Of course, for the relationship to be mutualistic it is essential that the animal digests only the fleshy fruit and not the seeds, which must remain viable when regurgitated or defecated. Thick, strong defenses that protect plant embryos are usually part of the price paid by the plant for dispersal by fruit-eaters. The plant kingdom has exploited a splendid array of morphological variations in the evolution of fleshy fruits (Figure 13.7).

Mutualisms involving animals that eat fleshy fruits and disperse seeds are seldom very specific to the species of animal concerned. Partly, this is because these mutualisms usually involve long-lived birds or mammals, and even in the tropics there are few plant species that fruit throughout the year and form a reliable food supply for any one specialist. But also, as will be apparent when pollination mutualisms are considered next, a more exclusive mutualistic link would require the plant's reward to be protected and denied to other animal species: this is much easier for nectar than for fruit. In any case, specialization by the animal is important in pollination, because interspecies transfers of pollen are disadvantageous, whereas with fruit and seed it is necessary only that they are dispersed away from the parent plant.

13.4.2 Pollination mutualisms

Most animal-pollinated flowers offer nectar, pollen or both as a reward to their visitors. Floral nectar seems to have no value to the plant other than as an attractant to animals and it has a cost to the plant, because the nectar carbohydrates might have been used in growth or some other activity.

Presumably, the evolution of specialized flowers and the involvement of animal pollinators have been favored because an animal may be able to recognize and discriminate between different flowers and so move pollen between different flowers of the same species but not to flowers of other species. Passive transfer of pollen, for example by wind or water, does not discriminate in this way and is therefore much more wasteful. Indeed, where the vectors and flowers are highly specialized, as is the case in many orchids, virtually no pollen is wasted even on the flowers of other species.

There are, though, costs that arise from adopting animals as mutualists in flower pollination. For example, animals carrying pollen may be responsible for the transmission of sexual diseases as well (Shykoff & Bucheli, 1995). The fungal pathogen *Microbotryum violaceum*, for example, is transmitted by pollinating visitors to the

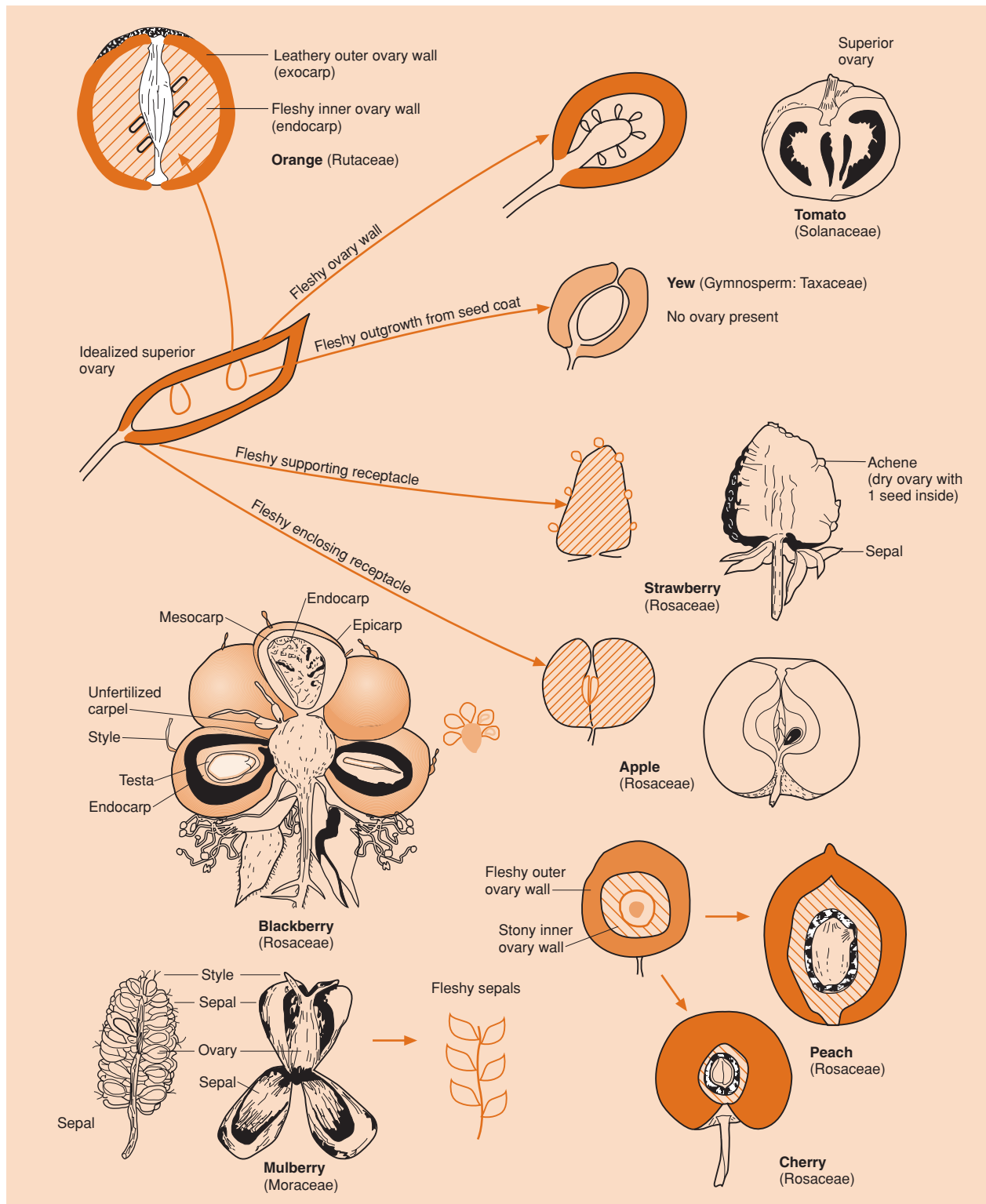


Figure 13.7 A variety of fleshy fruits involved in seed dispersal mutualisms illustrating morphological specializations that have been involved in the evolution of attractive fleshy structures.



Figure 13.8 Pollinators: (a) honeybee (*Apis mellifera*) on raspberry flowers, and (b) Cape sugarbird (*Promerops cafer*) feeding on *Protea eximia*. (Courtesy of Heather Angel.)

flowers of white campion (*Silene alba*) and in infected plants the anthers are filled with fungal spores.

**insect pollinators:
from generalists to
ultraspecialists**

Many different kinds of animals have entered into pollination liaisons with flowering plants, including hummingbirds, bats and even small rodents and marsupials (Figure 13.8). However,

the pollinators *par excellence* are, without doubt, the insects. Pollen is a nutritionally rich food resource, and in the simplest insect-pollinated flowers, pollen is offered in abundance and freely exposed to all and sundry. The plants rely for pollination on the insects being less than wholly efficient in their pollen consumption, carrying their spilt food with them from plant to plant. In more complex flowers, nectar (a solution of sugars) is produced as an additional or alternative reward. In the simplest of these, the nectaries are unprotected, but with increasing specialization the nectaries are enclosed in structures that restrict access to the nectar to just a few species of visitor. This range can be seen within the family Ranunculaceae. In the simple flower of *Ranunculus ficaria* the nectaries are exposed to all visitors, but in the more specialized flower of *R. bulbosus* there is a flap over the nectary, and in *Aquilegia* the nectaries have developed into long tubes and only visitors with long probosces (tongues) can reach the nectar. In the related *Aconitum* the whole flower is structured so that the nectaries are accessible only to insects of the right shape and size that are forced to brush against the anthers and pick up pollen. Unprotected nectaries have the advantage of a ready supply of pollinators, but because these pollinators are unspecialized they transfer much of the pollen to

the flowers of other species (though in practice, many generalists are actually 'sequential specialists', foraging preferentially on one plant species for hours or days). Protected nectaries have the advantage of efficient transfer of pollen by specialists to other flowers of the same species, but are reliant on there being sufficient numbers of these specialists.

Charles Darwin (1859) recognized that a long nectary, as in *Aquilegia*, forced a pollinating insect into close contact with the pollen at the nectary's mouth. Natural selection may then favor even longer nectaries, and as an evolutionary reaction, the tongues of the pollinator would be selected for increasing length – a reciprocal and escalating process of specialization. Nilsson (1988) deliberately shortened the nectary tubes of the long-tubed orchid *Platanthera* and showed that the flowers then produced many fewer seeds – presumably because the pollinator was not forced into a position that maximized the efficiency of pollination.

Flowering is a seasonal event in most plants, and this places strict **seasonality** limits on the degree to which a pollinator can become an obligate specialist. A pollinator can only become completely dependent on specific flowers as a source of food if its life cycle matches the flowering season of the plant. This is feasible for many short-lived insects like butterflies and moths, but longer lived pollinators such as bats and rodents, or bees with their long-lived colonies, are more likely to be generalists, turning from one relatively unspecialized flower to another through the seasons or to quite different foods when nectar is unavailable.



Figure 13.9 Fig wasps on a developing fig. Reproduced by permission of Gregory Dimijian/Science Photo Library.

13.4.3 Brood site pollination: figs and yuccas

figs and fig wasps . . .

Not every insect-pollinated plant provides its pollinator with only a take-away meal. In a number of cases, the plants also provide a home and sufficient food for the development of the insect larvae (Proctor *et al.*, 1996). The best studied of these are the complex, largely species-specific interactions between figs (*Ficus*) and fig wasps (Figure 13.9) (Wiebes, 1979; Bronstein, 1988). Figs bear many tiny flowers on a swollen receptacle with a narrow opening to the outside; the receptacle then becomes the fleshy fruit. The best-known species is the edible fig, *Ficus carica*. Some cultivated forms are entirely female and require no pollination for fruit to develop, but in wild *F. carica* three types of receptacle are produced at different times of the year. (Other species are less complicated, but the life cycle is similar.) In winter, the flowers are mostly neuter (sterile female) with a few male flowers near the opening. Tiny females of the wasp *Blastophaga psenes* invade the receptacle, lay eggs in the neuter flowers and then die. Each wasp larva then completes its development in the ovary of one flower, but the males hatch first and chew open the seeds occupied by the females and then mate with them. In early summer the females emerge, receiving pollen at the entrance from the male flowers, which have only just opened.

The fertilized females carry the pollen to a second type of receptacle, containing neuter and female flowers, where they lay their eggs. Neuter flowers, which cannot set seed, have a short style: the wasps can reach to lay their eggs in the ovaries where they develop. Female flowers, though, have long styles so the wasps cannot reach the ovaries and their eggs fail to develop, but in laying these eggs they fertilize the flowers, which set seed. Hence, these receptacles generate a combination of viable seeds (that benefit the fig) and adult fig wasps (that obviously benefit the wasps, but also benefit the figs since they are the figs' pollinators).

Following another round of wasp development, fertilized females emerge in the fall, and a variety of other animals eat the fruit and disperse the seeds. The fall-emerging wasps lay their eggs in a third kind of receptacle containing only neuter flowers, from which wasps emerge in winter to start the cycle again.

This, then, apart from being a fascinating piece of natural history, is a good example of a mutualism in which the interests of the two participants

. . . show mutualism despite conflict

none the less appear not to coincide. Specifically, the optimal proportion of flowers that develop into fig seeds and fig wasps is different for the two parties, and we might reasonably expect to see a negative correlation between the two: seeds produced *at the expense* of wasps, and vice versa (Herre & West, 1997). In fact, detecting this negative correlation, and hence establishing the conflict of interest, has proved elusive for reasons that frequently apply in studies of evolutionary ecology. The two variables tend, rather, to be *positively* correlated, since both tend to increase with two 'confounding' variables: the overall size of fruit and the overall proportion of flowers in a fruit that are visited by wasps. Herre and West (1997), however, in analyzing data from nine species of New World figs, were able to overcome this in a way that is generally applicable in such situations. They controlled statistically for variation in the confounding variables (asking, in effect, what the relationship between seed and wasp numbers would be in a fruit of constant size in which a constant proportion of flowers was visited) and then were able to uncover a negative correlation. The fig and fig wasp mutualists *do* appear to be involved in an on-going evolutionary battle.

A similar, and similarly much studied, set of mutualisms occurs between the 35–50 species of *Yucca* plant that live in North and Central

yuccas and yucca moths

America and the 17 species of yucca moth, 13 of which are newly described since 1999 (Pellmyr & Leebens-Mack, 2000). A female moth uses specialized 'tentacles' to collect together pollen from several anthers in one flower, which she then takes to the flower of another inflorescence (promoting outbreeding) where she both lays eggs in the ovaries and carefully deposits the pollen, again using her tentacles. The development of the moth larvae requires successful pollination, since unpollinated flowers quickly die, but the larvae also consume seeds in their immediate vicinity, though many other seeds develop successfully. On completing their development, the larvae drop to the soil to pupate, emerging one or more years later during the yucca's flowering season. The reproductive success of an individual adult female moth is not, therefore, linked to that of an individual yucca plant in the same way as are those of female fig wasps and figs.

A detailed review of both seed dispersal and pollination mutualisms is given by Thompson (1995), who provides a thorough account of the processes that may lead to the evolution of such mutualisms.

13.5 Mutualisms involving gut inhabitants

Most of the mutualisms discussed so far have depended on patterns of behavior, where neither species lives entirely 'within' its partner. In many other mutualisms, one of the partners is a unicellular eukaryote or bacterium that is integrated more or less permanently into the body cavity or even the cells of its multicellular partner. The microbiota occupying parts of various animals' alimentary canals are the best known extracellular symbionts.

13.5.1 Vertebrate guts

The crucial role of microbes in the digestion of cellulose by vertebrate herbivores has long been appreciated, but it now appears that the gastrointestinal tracts of all vertebrates are populated by a mutualistic microbiota (reviewed in Stevens & Hume, 1998). Protozoa and fungi are usually present but the major

contributors to these 'fermentation' processes are bacteria. Their diversity is greatest in regions of the gut where the pH is relatively neutral and food retention times are relatively long. In small mammals (e.g. rodents, rabbits and hares) the cecum is the main fermentation chamber, whereas in larger nonruminant mammals such as horses the colon is the main site, as it is in elephants, which, like rabbits, practice coprophagy (consume their own feces) (Figure 13.10). In ruminants, like cattle and sheep, and in kangaroos and other marsupials, fermentation occurs in specialized stomachs.

The basis of the mutualism is straightforward. The microbes receive a steady flow of substrates for growth in the form of food that has been eaten, chewed and partly homogenized. They live within a chamber in which pH and, in endotherms, temperature are regulated and anaerobic conditions are maintained. The vertebrate hosts, especially the herbivores, receive nutrition from food that they would otherwise find, literally, indigestible. The bacteria produce short-chain fatty acids (SCFAs) by fermentation of the host's dietary cellulose and starches and of the endogenous

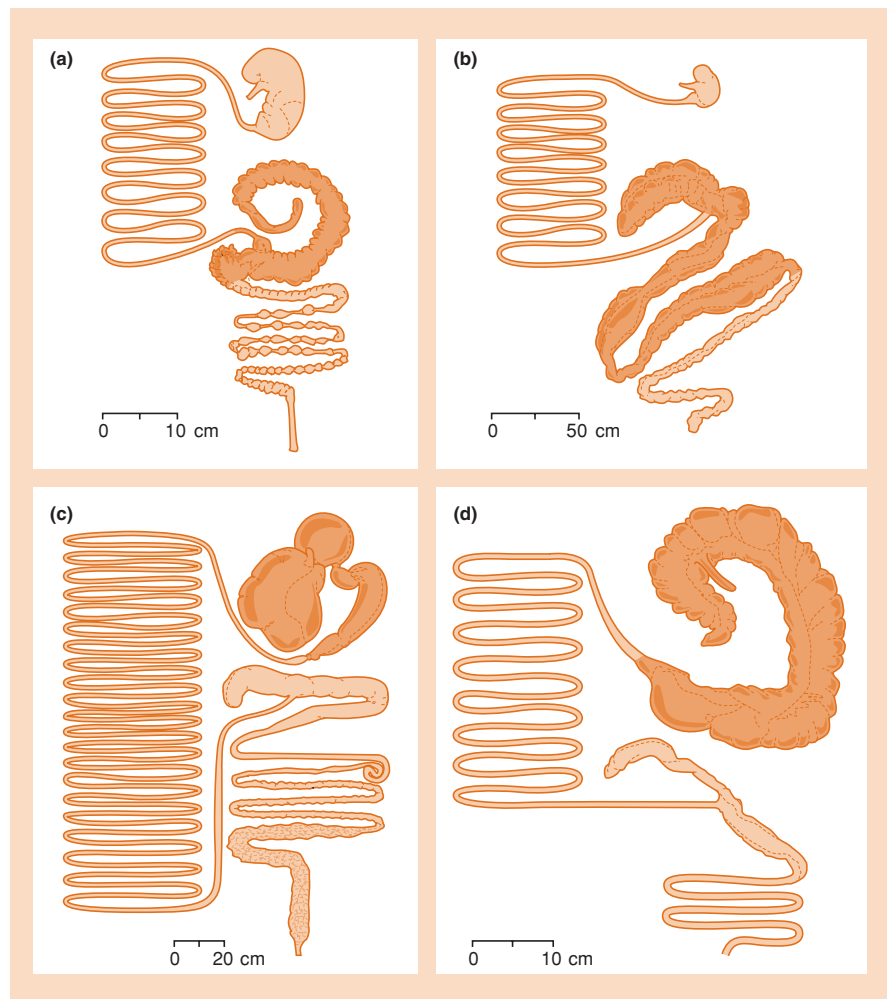


Figure 13.10 The digestive tracts of herbivorous mammals are commonly modified to provide fermentation chambers inhabited by a rich fauna and flora or microbes. (a) A rabbit, with a fermentation chamber in the expanded cecum. (b) A zebra, with fermentation chambers in both the cecum and colon. (c) A sheep, with foregut fermentation in an enlarged portion of the stomach, rumen and reticulum. (d) A kangaroo, with an elongate fermentation chamber in the proximal portion of the stomach. (After Stevens & Hume, 1998.)

Species	Function	Products
<i>Bacteroides succinogenes</i>	C, A	F, A, S
<i>Ruminococcus albus</i>	C, X	F, A, E, H, C
<i>R. flavefaciens</i>	C, X	F, A, S, H
<i>Butyrivibrio fibrisolvens</i>	C, X, PR	F, A, L, B, E, H, C
<i>Clostridium lochheadii</i>	C, PR	F, A, B, E, H, C
<i>Streptococcus bovis</i>	A, SS, PR	L, A, F
<i>B. amylophilus</i>	A, P, PR	F, A, S
<i>B. ruminicola</i>	A, X, P, PR	F, A, P, S
<i>Succinimonas amylolytica</i>	A, D	A, S
<i>Selenomonas ruminantium</i>	A, SS, GU, LU, PR	A, L, P, H, C
<i>Lachnospira multiparus</i>	P, PR, A	F, A, E, L, H, C
<i>Succinivibrio dextrinosolvans</i>	P, D	F, A, L, S
<i>Methanobrevibacter ruminantium</i>	M, HU	M
<i>Methanosarcina barkeri</i>	M, HU	M, C
<i>Spirochete species</i>	P, SS	F, A, L, S, E
<i>Megasphaera elsdenii</i>	SS, LU	A, P, B, V, CP, H, C
<i>Lactobacillus sp.</i>	SS	L
<i>Anaerovibrio lipolytica</i>	L, GU	A, P, S
<i>Eubacterium ruminantium</i>	SS	F, A, B, C

Functions: A, amylolytic; C, cellulolytic; D, dextrinolytic; GU, glycerol utilizing; HU, hydrogen utilizer; L, lipolytic; LU, lactate utilizing; M, methanogenic; P, pectinolytic; PR, proteolytic; SS, major soluble sugar fermenter; X, xylanolytic.

Products: A, acetate; B, butyrate; C, carbon dioxide; CP, caproate; E, ethanol; F, formate; H, hydrogen; L, lactate; M, methane P, propionate; S, succinate; V, valerate;.

Table 13.1 A number of the bacterial species of the rumen, illustrating their wide range of functions and the wide range of products that they generate. (After Allison, 1984; Stevens & Hume, 1998.)

carbohydrates contained in host mucus and sloughed epithelial cells. SCFAs are often a major source of energy for the host; for example, they provide more than 60% of the maintenance energy requirements for cattle and 29–79% of those for sheep (Stevens & Hume, 1998). The microbes also convert nitrogenous compounds (amino acids that escape absorption in the midgut, urea that would otherwise be excreted by the host, mucus and sloughed cells) into ammonia and microbial protein, conserving nitrogen and water; and they synthesize B vitamins. The microbial protein is useful to the host if it can be digested – in the intestine by foregut fermenters and following coprophagy in hindgut fermenters – but ammonia is usually not useful and may even be toxic to the host.

13.5.2 Ruminant guts

The stomach of ruminants comprises a three-part forestomach (rumen, reticulum and omasum) followed by an enzyme-secreting abomasum that is similar to the whole stomach of most other vertebrates. The rumen and reticulum are the main sites of fermentation, and the omasum serves largely to transfer material to the abomasum. Only particles with a volume of about 5 μl or less can pass from the reticulum into the omasum; the animal regurgitates and rechews the larger particles (the pro-

cess of rumination). Dense populations of bacteria (10^{10} – 10^{11} ml^{-1}) and protozoa (10^5 – 10^6 ml^{-1} but occupying a similar volume to the bacteria) are present in the rumen. The bacterial communities of the rumen are composed almost wholly of obligate anaerobes – many are killed instantly by exposure to oxygen – but they perform a wide variety of functions (subsist on a wide variety of substrates) and generate a wide range of products (Table 13.1). Cellulose and other fibers are the important constituents of the ruminant's diet, and the ruminant itself lacks the enzymes to digest these. The cellulolytic activities of the rumen microflora are therefore of crucial importance. But not all the bacteria are cellulolytic: many subsist on substrates (lactate, hydrogen) generated by other bacteria in the rumen.

The protozoa in the gut are also a complex mixture of specialists. Most are holotrich ciliates and entodiniomorphs. A few can digest cellulose. The cellulolytic ciliates have intrinsic cellulases, although some other protozoa may use bacterial symbionts. Some consume bacteria: in their absence the number of bacteria rise. Some of the entodiniomorphs prey on other protozoa. Thus, the diverse processes of competition, predation and mutualism, and the food chains that characterize terrestrial and aquatic communities in nature, are all present within the rumen microcosm.

a complex
community of
mutualists

13.5.3 Refection

Eating feces is a taboo amongst humans, presumably through some combination of biological and cultural evolution in response to the health hazards posed by pathogenic microbes, including many that are relatively harmless in the hindgut but are pathogenic in more anterior regions. For many vertebrates, however, symbiotic microbes, living in the hindgut beyond the regions where effective nutrient absorption is possible, are a resource that is too good to waste. Thus coprophagy (eating feces) or refection (eating one's own feces) is a regular practice in many small herbivorous mammals. This is developed to a fine art in species such as rabbits that have a 'colonic separation mechanism' that allows them to produce separate dry, non-nutritious fecal pellets and soft, more nutritious pellets that they consume selectively. These contain high levels of SCFAs, microbial protein and B vitamins, and can provide 30% of a rabbit's nitrogen requirements and more B vitamins than it requires (Björnhag, 1994; Stevens & Hume, 1998).

13.5.4 Termite guts

Termites are social insects of the order Isoptera, many of

which depend on mutualists for the digestion of wood. Primitive termites feed directly on wood, and most of the cellulose, hemicelluloses and possibly lignins are digested by mutualists in the gut (Figure 13.11), where the paunch (part of the segmented hindgut) forms a microbial fermentation chamber. However, the advanced termites (75% of all the species) rely much more heavily on their own cellulase (Hogan *et al.*, 1988), while a third group (the Macrotermitinae) cultivate wood-digesting fungi that the termites eat along with the wood itself, which the fungal cellulases assist in digesting.

Termites redefecate, so that food material passes at least twice through the gut, and microbes that have reproduced during the first passage may be digested the second time round. The major group of microorganisms in the paunch of primitive termites are anaerobic flagellate protozoans. Bacteria are also present, but cannot digest cellulose. The protozoa engulf particles of wood and ferment the cellulose within their cells, releasing carbon dioxide and hydrogen. The principal products, subsequently absorbed by the host, are SCFAs (as in vertebrates) but in termites they are primarily acetic acid.

The bacterial population of the termite gut is less conspicuous than that of the rumen, but appears to play a part in two distinct mutualisms.

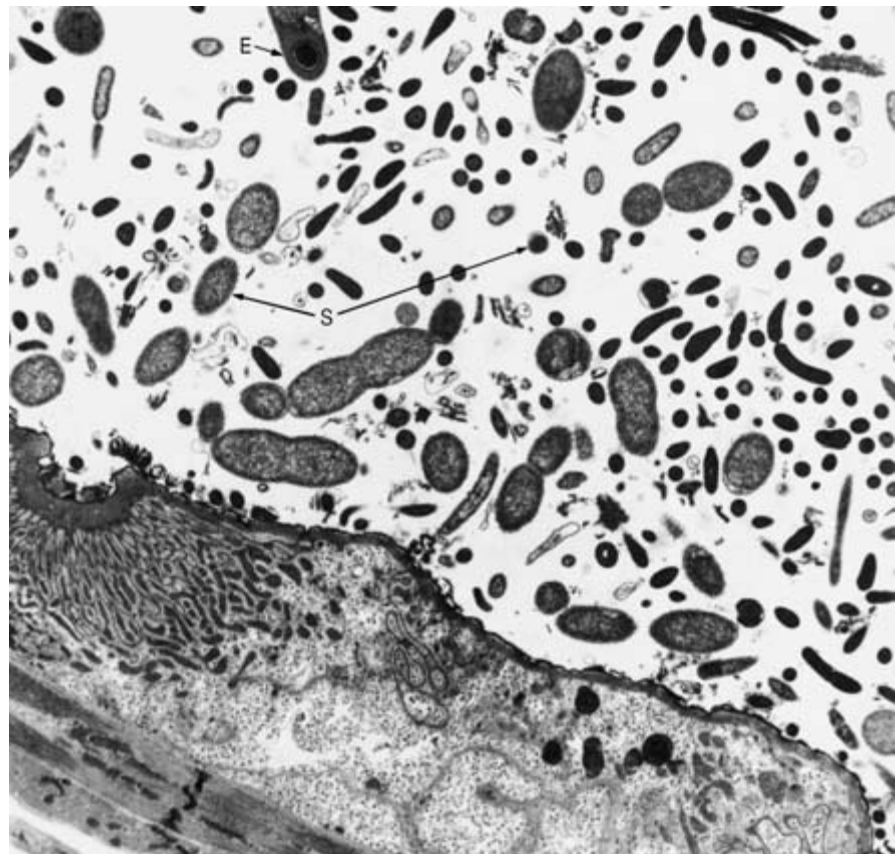


Figure 13.11 Electron micrograph of a thin section of the paunch of the termite *Reticulitermes flavipes*. Much of the flora is composed of aggregates of bacteria. Amongst them can be seen endospore-forming bacteria (E), spirochetes (S) and protozoa. (After Breznak, 1975.)

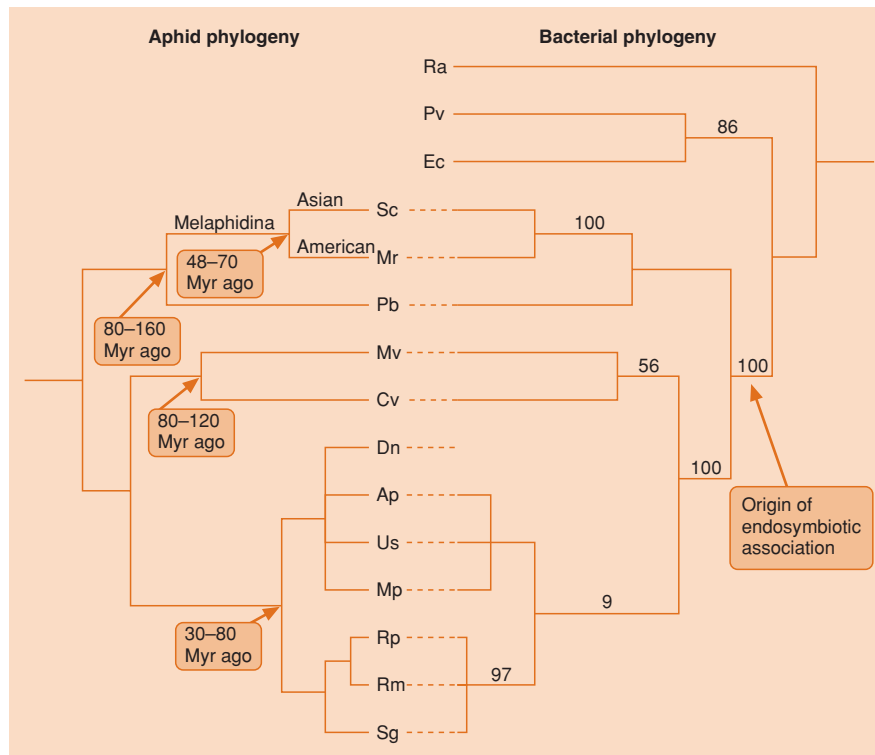


Figure 13.12 The phylogeny of selected aphids and their corresponding primary endosymbionts. Other bacteria are shown for comparison. The aphid phylogeny (after Heie, 1987) is shown on the left and the bacterial phylogeny on the right. Broken lines connect the associated aphids and bacteria. Three species of bacteria that are not endosymbionts are also shown in the phylogeny: Ec, *Escherichia coli*; Pv, *Proteus vulgaris*; Ra, *Ruminobacter amylophilus* (a rumen symbiont). The distances along the branches are drawn to be roughly proportionate to time. (After Moran *et al.*, 1993.) Aphid species: Ap, *Acyrtosiphon pisum*; Cv, *Chaitophorus viminalis*; Dn, *Diuraphis noxia*; Mp, *Myzus persicae*; Mr, *Melaphis rhois*; Mv, *Mindarus victorariae*; Pb, *Pemphigus betae*; Rm, *Rhopalosiphum maidis*; Rp, *Rhodolophium padi*; Sc, *Schlectendalia chinensis*; Sg, *Schizaphis graminum*; Us, *Uroleucon sonchi*.

- 1 Spirochetes tend to be concentrated at the surface of the flagellates. The spirochetes possibly receive nutrients from the flagellates, and the flagellates gain mobility from the movements of the spirochetes: a pair of mutualists living mutualistically within a third species.
- 2 Some bacteria in the termite gut are capable of fixing gaseous nitrogen – apparently the only clearly established example of nitrogen-fixing symbionts in insects (Douglas, 1992). Nitrogen fixation stops when antibacterial antibiotics are eaten (Breznak, 1975), and the rate of nitrogen fixation falls off sharply if the nitrogen content of the diet is increased.

13.6 Mutualism within animal cells: insect mycetocyte symbioses

In mycetocyte symbioses between microorganisms and insects, the maternally inherited microorganisms are found within the cytoplasm of specialized cells, mycetocytes, and the interaction is unquestionably mutualistic. It is required by the insects for the nutritional benefits the microorganisms bring, as key providers of essential amino acids, lipids and vitamins, and is required by the microorganisms for their very existence (Douglas, 1998). The symbioses are found in a wide variety of types of insect, and are universally or near-universally present in cockroaches, homopterans, bed bugs, sucking lice, tsetse flies, lyctid beetles and

camponotid ants. They have evolved independently in different groups of microorganisms and their insect partners, but in effectively all cases the insects live their lives on nutritionally poor or unbalanced diets: phloem sap, vertebrate blood, wood and so on. Mostly the symbionts are various sorts of bacteria, although in some insects yeasts are involved.

Amongst these symbioses, most is known by far about the interactions between aphids and bacteria in the genus *Buchnera* (Douglas, 1998). The mycetocytes are found in the hemocoel of the aphids and the bacteria occupy around 60% of the mycetocyte cytoplasm. The bacteria cannot be brought into culture in the laboratory and have never been found other than in aphid mycetocytes, but the extent and nature of the benefit they bring to the aphids can be studied by removing the *Buchnera* by treating the aphids with antibiotics. Such ‘aposymbiotic’ aphids grow very slowly and develop into adults that produce few or no offspring. The most fundamental function performed by the bacteria is to produce essential amino acids that are absent in phloem sap from nonessential amino acids like glutamate, and antibiotic treatment confirms that the aphids cannot do this alone. In addition, though, the *Buchnera* seem to provide other benefits, since symbiotic aphids still outperform aposymbiotic aphids when the latter are provided with all the essential amino acids, but establishing further nutritional functions has proved elusive.

aphids and
Buchnera . . .

... provide an ecological and evolutionary link

The aphid–*Buchnera* interaction also provides an excellent example of how an intimate association between mutualists may link them at both the ecological and the evolutionary level. The

Buchnera are transmitted transovarially, that is, they are passed by a mother to her offspring in her eggs. Hence, an aphid lineage supports a corresponding single *Buchnera* lineage, and this is no doubt the reason for the strictly congruent phylogenies of aphid and *Buchnera* species: each aphid species has its own *Buchnera* species (see, for example, Figure 13.12). Moreover, these molecular studies, which allow the *Buchnera* phylogeny to be reconstructed, also suggest that the aphids acquired *Buchnera* just once in their evolutionary history, apparently between 160 and 280 million years ago, after the divergence from the main aphid lineage of the only two aphid families not to have a mycetocyte symbiosis, the phylloxerids and the adelgids (Moran *et al.*, 1993). Providing a final twist, the only other aphids without *Buchnera* (in the family Hormaphididae) appear to have lost them secondarily in their evolutionary history, but they do instead host symbiotic yeasts (Douglas, 1998). It seems more likely that the yeasts competitively displaced the bacteria than that the bacteria were first lost and the yeasts subsequently acquired.

Lastly, Douglas (1998) also points out that whereas all Homoptera that feed on nutritionally deficient phloem sap have mycetocyte symbioses, including the aphids described above, those that have switched secondarily in their evolutionary history to feeding on intact plant cells have lost the symbiosis. This, then, is an illustration from a comparative, evolutionary perspective that even in clearly mutualistic symbioses like these, the benefit is a *net* benefit. Once the insects' requirements are reduced, as in a switch of diet, the balance of the costs and benefits of the symbionts is also changed. In this case, the costs clearly outweigh the benefits on a changed diet: those insects that *lost* their symbionts have been favored by natural selection.

13.7 Photosynthetic symbionts within aquatic invertebrates

Hydra and Chlorella

Algae are found within the tissues of a variety of animals, particularly in the phylum Cnidaria. In freshwater symbioses the algal symbiont is usually *Chlorella*. For example, in *Hydra viridis*, cells of *Chlorella* are present in large numbers (1.5×10^5 per hydroid) within the digestive cells of the endoderm. In the light, a *Hydra* receives photosynthates from the algae and 50–100% of its oxygen needs. It can also use organic food. Yet when a *Hydra* is maintained in darkness and fed daily with organic food, a reduced symbiotic population of algae is maintained for at least 6 months that can return to normal within 2 days of exposure to light (Muscatine & Pool, 1979). Thus,

armed with its symbionts, and depending on local conditions and resources, *Hydra* can behave both as an autotroph and a heterotroph. There must then be regulatory processes harmonizing the growth of the endosymbiont and its host (Douglas & Smith, 1984), as there must presumably be in all such symbioses. If this were not the case, the symbionts would either overgrow and kill the host or fail to keep pace and become diluted as the host grew.

There are many records of close associations between algae and protozoa in the marine plankton. For example, in the ciliate *Mesodinium rubrum*, 'chloroplasts' are present that appear to be symbiotic algae. The mutualistic consortium of protists and algae can fix carbon dioxide and take up mineral nutrients, and often forms dense populations known as 'red tides' (e.g. Crawford *et al.*, 1997). Extraordinarily high production rates have been recorded from such populations (in excess of $2 \text{ g m}^{-3} \text{ h}^{-1}$ of carbon) – apparently the highest levels of primary productivity ever recorded for populations of aquatic microorganisms.

marine plankton

13.7.1 Reef-building corals and coral bleaching

We have already noted that mutualists dominate environments around the world in terms of their biomass. Coral reefs provide an important example: reef-building corals (another dramatic example of autogenic ecosystem engineering – see Section 13.1) are in fact mutualistic associations between heterotrophic Cnidaria and phototrophic dinoflagellate algae from the genus *Symbiodinium*. Coral reefs provide an illustration, too, of the potential vulnerability of even the most dominant of 'engineered' habitat features. There have been repeated reports of 'coral bleaching' since it was first described in 1984: the whitening of corals as a result of the loss of the endosymbionts and/or their photosynthetic pigments (Brown, 1997). Bleaching occurs mainly in response to unusually elevated temperatures (as seen at the Phuket study site, Thailand; Figure 13.13a), but also in response to high intensities of solar radiation and even disease. Thus, episodes of bleaching seem likely to become increasingly frequent as global temperatures rise (Figure 13.13a; see Section 2.8.2), which is a particular cause for concern, since some bleaching episodes have been followed by mass mortality of corals. This was apparent at Phuket, for example, associated with the bleaching episodes of 1991 and 1995 (Figure 13.13b). (On the other hand, a more catastrophic loss had occurred in 1987 as a result, not of bleaching, but of dredging activity, and the decline in cover in the early 1990s appeared to result from an *interaction* between bleaching and a variety of local human disturbances.)

We clearly cannot be complacent about the effects of global warming on coral reefs – and there *are* likely always

bleaching and global warming

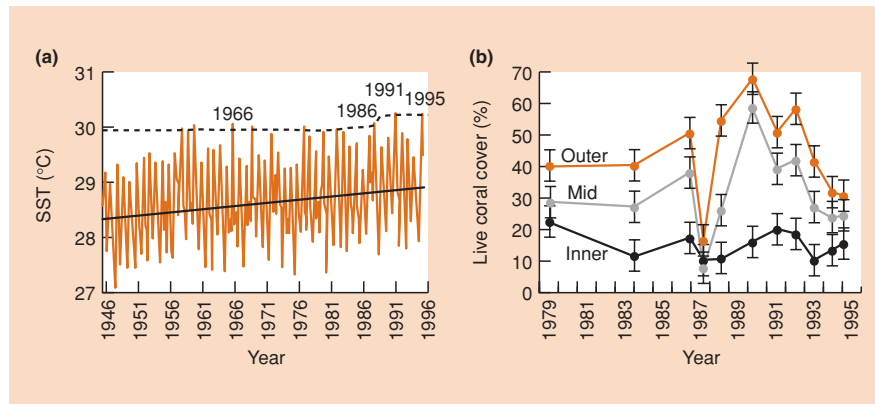


Figure 13.13 (a) Monthly mean sea surface temperatures (SSTs) for sea areas off Phuket, Thailand, from 1945 to 1995. The regression line for all points is shown ($P < 0.001$). The dashed line drawn at 30.11°C represents a tentative bleaching threshold. The years exceeding this are shown: bleaching was observed in 1991 and 1995 but not monitored prior to that. (b) Mean percentage coral cover (\pm SE) on inner (—), mid (—) and outer (—) reef flats at Phuket, Thailand, over the period 1979–95. (After Brown, 1997.)

to be human disturbances with which bleaching effects can interact – but it is also apparent that reef corals are able to acclimate to the changed conditions that may induce bleaching and to recover from bleaching episodes. Their adaptability is illustrated by another study at Phuket. During the 1995 episode, it had been observed that bleaching in the coral *Goniastrea aspera* occurred predominantly on east- rather than west-facing surfaces. The latter normally suffer greater exposure to solar radiation, which also has a tendency to cause bleaching. This therefore suggests that tolerance to bleaching had been built up in the west-facing corals. Such a difference in tolerance was confirmed experimentally (Figure 13.14): there was little or no bleaching on the ‘adapted’ west-facing surfaces at high temperatures.

another mutualism
extending beyond
two species

Meanwhile, another study of coral bleaching adds to the growing realization that seemingly simple two-species mutualisms may be more complex and subtler than might be imagined. The ecologically dominant Caribbean corals *Montastraea annularis* and *M. faveolata* both host three quite separate ‘species’ or ‘phylotypes’ of *Symbiodinium* (denoted *A*, *B* and *C* and distinguishable only by genetic methods). Phylotypes *A* and *B* are common in shallower, high-irradiance habitats, whereas *C* predominates in deeper, lower irradiance sites – illustrated both by comparisons of colonies from different depths and of samples from different depths within a colony (Figure 13.14b). In the fall of 1995, following a prolonged period above the mean maximum summer temperature, bleaching occurred in *M. annularis* and *M. faveolata* in the reefs off Panama and elsewhere. Bleaching, however, was rare at the shallowest and the deepest sites, but was most apparent in shallower colonies at shaded sites and in deeper colonies at more exposed sites. A comparison of adjacent samples before and after bleaching provides an explanation (Figure 13.14c). The bleaching resulted from the selective loss of *Symbiodinium C*. It appears to have occurred at locations supporting *C* and one or both of the other two species, near the irradiance limit of *C* under non-

bleaching conditions. At shaded deep-water sites, dominated by *C*, the high temperatures in 1995 were not sufficient to push *C* into bleaching conditions. The shallowest sites were occupied by the species *A* and *B*, which were not susceptible to bleaching at these temperatures. Bleaching occurred, however, where *C* was initially present but was pushed beyond its limit by the increased temperature. At these sites, the loss of *C* was typically close to 100%, *B* decreased by around 14%, but *A* more than doubled in three of five instances.

It seems, therefore, first, that the coral–*Symbiodinium* mutualism involves a range of endosymbionts that allows the corals to thrive in a wider range of habitats than would otherwise be possible. Second, looking at the mutualism from the algal side, the endosymbionts must constantly be engaged in a competitive battle, the balance of which alters over space and time (see Section 8.5). Finally, bleaching (and subsequent recovery), and possibly also ‘adaptation’ of the type described above, may be seen as manifestations of this competitive battle: not breakdowns and reconstructions in a simple two-species association, but shifts in a complex symbiotic community.

13.8 Mutualisms involving higher plants and fungi

A wide variety of symbiotic associations are formed between higher plants and fungi. A very remarkable group of Ascomycete fungi, the Clavicipitaceae, grow in the tissues of many species of grass and a few species of sedge. The family includes species that are easily recognized as parasites (e.g. *Claviceps*, the ergot fungus, and *Epichloe*, the choke disease of grasses), others that are clearly mutualistic, and a large number where the costs and benefits are uncertain. The fungal mycelia characteristically grow as sparsely branched filaments running through intercellular spaces along the axis of leaves and stems, but they are not found in roots. Many of the symbiotic fungi produce powerful toxic alkaloids that

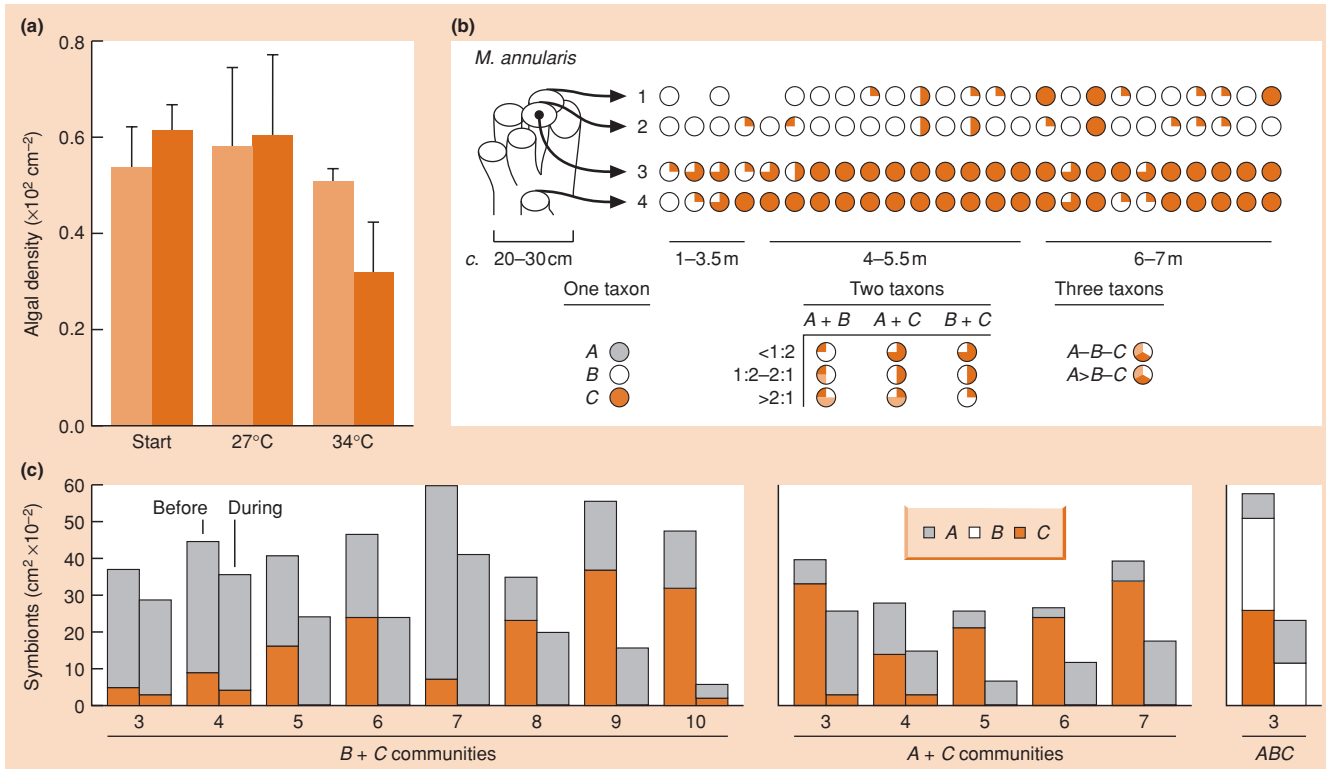


Figure 13.14 Coral acclimation and recovery in coral bleaching. (a) Algal density in western (light bars) and eastern (dark bars) cores of the coral *Goniastrea aspera* before and after exposure to elevated (34°C) and ambient (27°C) temperatures for 68 h. Mean values are shown; error bars show 1 SD ($n = 5$). (After Brown *et al.*, 2000.) (b) Symbiont communities in another coral, *Montastraea annularis*, collected in January 1995 off the coast of Panama. Each symbol represents a sample that contained the algal taxa *Symbiodinium A*, *B* or *C*, or mixtures of taxa summarized according to the code shown below. Columns in the data represent individual coral colonies (depth increases from left to right) and rows represent locations of higher (rows 1 and 2) and lower (rows 3 and 4) irradiance, as defined in the diagram to the left. (After Rowan *et al.*, 1997.) (c) Corresponding symbiont communities from close to the bleaching region of *Symbiodinium C* before (January 1995) and during (October 1995) an episode of coral bleaching. Densities of *A* (gray), *B* (white) and *C* (orange) before and during bleaching (left and right bars of each pair, respectively) in samples reported in *B + C* communities (3–10), *A + C* communities (3–7) and an *ABC* community. (After Rowan *et al.*, 1997.)

confer some protection from grazing animals (the evidence is reviewed in Clay, 1990) and, perhaps even more important, deter seed predators (Knoch *et al.*, 1993).

**not roots but
mycorrhizas**

A quite different mutualism of fungi with higher plants occurs in roots. Most higher plants do not have roots, they have mycorrhizas – intimate mutualisms between fungi and root tissue. Plants of only a few families like the Cruciferae are the exception. Broadly, the fungal networks in mycorrhizas capture nutrients from the soil, which they transport to the plants in exchange for carbon. Many plant species can live without their mycorrhizal fungi in soils where neither nutrients nor water are ever limiting, but in the harsh world of natural plant communities, the symbioses, if not strictly obligate, are none the less ‘ecologically obligate’. That is,

they are necessary if the individuals are to survive in nature (Buscot *et al.*, 2000). The fossil record suggests that the earliest land plants, too, were heavily infected. These species lacked root hairs, even roots in some cases, and the early colonization of the land may have depended on the presence of the fungi to make the necessary intimate contact between plants and substrates.

Generally, three major types of mycorrhiza are recognized. Arbuscular mycorrhizas are found in up to two-thirds of all plant species, including most nonwoody species and tropical trees. Ectomycorrhizal fungi form symbioses with many trees and shrubs, dominating boreal and temperate forests and also some tropical rainforests. Finally, ericoid mycorrhizas are found in the dominant species of heathlands including the northern hemisphere heaths and heathers (Ericaceae) and the Australian heaths (Epacridaceae).



Figure 13.15 Mycorrhiza of pine (*Pinus sylvestris*). The swollen, much branched structure is the modified rootlet enveloped in a thick sheath of fungal tissue. (Courtesy of J. Whiting; photograph by S. Barber.)

13.8.1 Ectomycorrhizas

An estimated 5000–6000 species of Basidiomycete and Ascomycete fungi form ectomycorrhizas (ECMs) on the roots of trees (Buscot *et al.*, 2000). Infected roots are usually concentrated in the litter layer of the soil. Fungi form a sheath or mantle of varying thickness around the roots. From there, hyphae radiate into the litter layer, extracting nutrients and water and also producing large fruiting bodies that release enormous numbers of wind-borne spores. The fungal mycelium also extends inwards from the sheath, penetrating between the cells of the root cortex to give intimate cell-to-cell contact with the host and establishing an interface with a large surface area for the exchange of photo-assimilates, soil water and nutrients between the host plant and its fungal partner. The fungus usually induces morphogenetic changes in the host roots, which cease to grow apically and remain stubby (Figure 13.15). Host roots that penetrate into the deeper, less organically rich layers of the soil continue to elongate.

The ECM fungi (see Buscot *et al.*, 2000 for a review) are effective in extracting the sparse and patchy supplies of phosphorus and especially nitrogen from the forest litter layer, and their high species diversity presumably reflects a corresponding diversity of niches in this environment (though this diversity of niches is very far from having been demonstrated). Carbon flows from the plant to the fungus, very largely in the form of the simple hexose sugars: glucose and fructose. Fungal consumption of these may represent up to 30% of the plants' net rate of photosynthate production. The plants, though, are often nitrogen-limited, since in the forest litter there are low rates of nitrogen mineralization (conversion from organic to inorganic forms), and inorganic

nitrogen is itself mostly available as ammonia. It is therefore crucial for forest trees that ECM fungi can access organic nitrogen directly through enzymic degradation, utilize ammonium as a preferred source of inorganic nitrogen, and circumvent ammonium depletion zones through extensive hyphal growth. None the less, the idea that this relationship between the fungi and their host plants is mutually exploitative rather than 'cosy' is emphasized by its responsiveness to changing circumstances. ECM growth is directly related to the rate of flow of hexose sugars from the plant. But when the direct availability of nitrate to the plants is high, either naturally or through artificial supplementation, plant metabolism is directed away from hexose production (and export) and towards amino acid synthesis. As a result the ECM degrades; the plants seem to support just as much ECM as they appear to need.

13.8.2 Arbuscular mycorrhizas

Arbuscular mycorrhizas (AMs) do not form a sheath but penetrate *within* the roots of the host, though they do not alter the host's root morphology. Roots become infected from mycelium present in the soil or from germ tubes that develop from asexual spores, which are very large and produced in small numbers – a striking contrast with the ECM fungi. Initially, the fungus grows between host cells but then enters them and forms a finely branched intracellular 'arbuscule'. The fungi responsible comprise a distinct phylum, the Glomeromycota (Schüßler *et al.*, 2001). Although originally divided into only about 150 species, suggesting a lack of host specificity (since there are vastly more species of hosts), modern genetic methods have uncovered a far greater

diversity among the AM fungi, and there is increasing evidence of niche differentiation amongst them. For instance, when 89 root samples were taken from three grass species that co-occurred in the same plots in a field experiment, and their AM fungi were characterized using such a method – terminal restriction fragment length polymorphism – there was clear separation amongst the AM strains found on the different hosts (Figure 13.16).

a range of benefits?

There has been a tendency to emphasize facilitation of the uptake of phosphorus as the main benefit to plants from AM symbioses (phosphorus is a highly immobile element in the soil, which is therefore frequently limiting to plant growth), but the truth appears to be more complex than this. Benefits have been demonstrated, too, in nitrogen uptake, pathogen and herbivore protection, and resistance to toxic metals (Newsham *et al.*, 1995). Certainly, there are cases where the inflow of phosphorus is strongly related to the degree of colonization of roots by AM fungi. This has been shown for the bluebell, *Hyacinthoides non-scripta*, as colonization progresses during its phase of subterranean growth from August to February through to its above-ground photosynthetic phase thereafter (Figure 13.17a). Indeed, bluebells cultured without AM fungi are unable to take up phosphorus through their poorly branched system of roots (Merryweather & Fitter, 1995).

On the other hand, a factorial set of experiments examined the growth of the annual grass *Vulpia ciliata* ssp. *ambigua* at sites

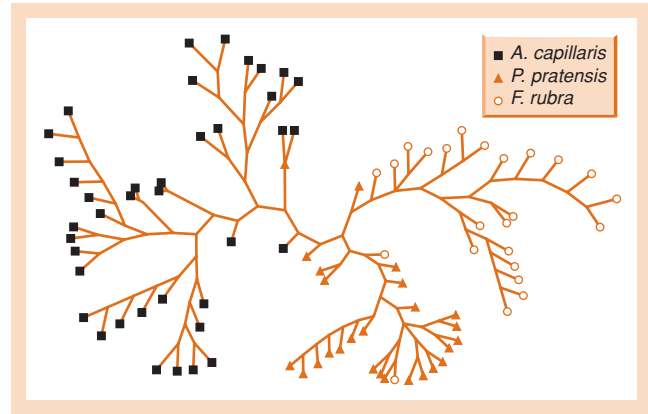


Figure 13.16 The similarity among 89 arbuscular mycorrhiza (AM) fungal communities taken from the roots of three coexisting grass species, *Agrostis capillaris*, *Poa pratensis* and *Festuca rubra*, assessed by terminal restriction fragment length polymorphism. Each terminal on the ‘tree’ is a different sample, with the grass species from which it originated shown. More similar samples are closer together on the tree. The similarity within, and the differentiation between, the AM fungal communities associated with different hosts are plainly apparent. (After Vandenkoornhuysen *et al.*, 2003.)

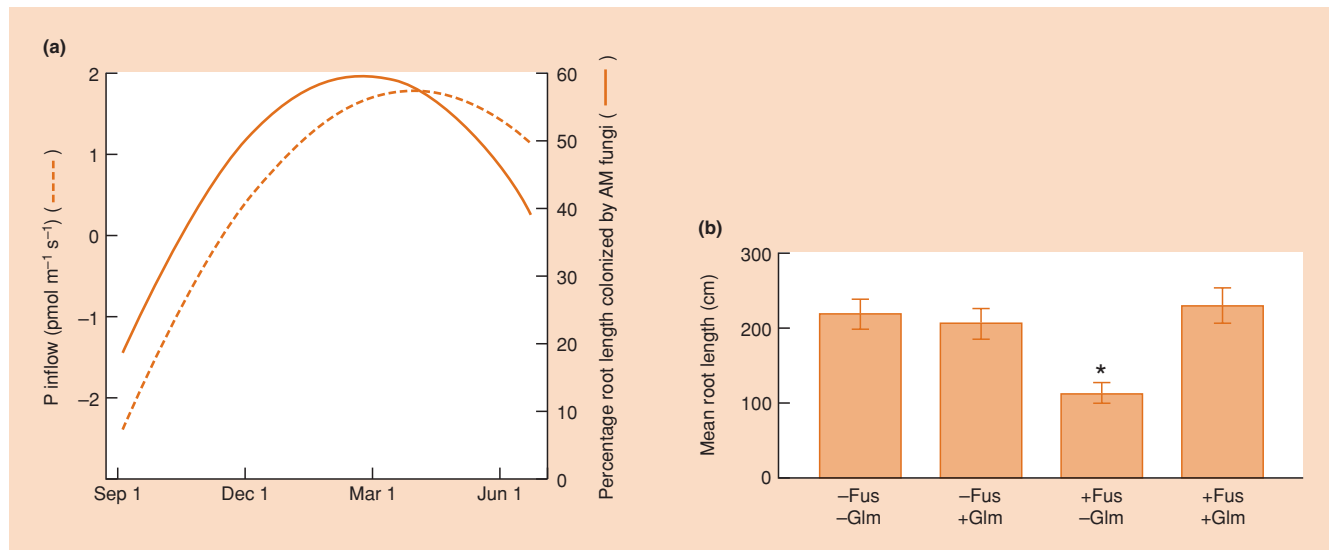


Figure 13.17 (a) Curves fitted to rates of phosphorus inflow (---, left axis) and root colonization by arbuscular mycorrhiza (AM) fungi (—, right axis) in the bluebell, *Hyacinthoides non-scripta*, over a single growing season. (After Merryweather & Fitter, 1995; Newsham *et al.*, 1995.) (b) The effects of a factorial combination of *Fusarium oxysporum* (Fus) and an AM fungus, *Glomus* sp. (Glm), on the growth (root length) of *Vulpia* plants. Values are means of 16 replicates per treatment; bars show standard errors; the asterisk signifies a significant difference at $P < 0.05$ in a Fisher’s pairwise comparison. (After Newsham *et al.*, 1994, 1995.)

in eastern England where there were large differences in the intensity of natural mycorrhizal infection (West *et al.*, 1993). In one treatment phosphate was applied, and in another the fungicide benomyl was used to control the fungal infection. Fecundity of the grass was scarcely affected by any of the treatments. An explanation was provided by a further set of experiments (Figure 13.17b) in which seedlings of *Vulpia* were grown with an AM fungus (*Glomus* sp.), with the pathogenic fungus *Fusarium oxysporum*, with both, and with neither. Growth was not enhanced by *Glomus* alone, but growth was harmed by *Fusarium* in the absence of *Glomus*. When both were present, growth returned to normal levels. Clearly, the mycorrhiza did not benefit the phosphorus-economy of the *Vulpia*, but it did protect it from the harmful effects of the pathogen. (In the previous experiment, benomyl presumably had no effect on performance because it controlled both mycorrhizal *and* pathogenic fungi.)

it depends on
the species

The key difference appears to be that *Vulpia*, unlike the bluebell, has a highly branched system of roots, and Newsham *et al.* (1995) go so far as to

propose a continuum of AM function in relation to root architecture, with *Vulpia* and *Hyacinthoides* sitting towards the two extremes. Plants with finely branched roots have little need for supplementary phosphorus capture, but development of that same root architecture provides multiple points of entry for plant pathogens. In such cases AM symbioses are therefore likely to have evolved with an emphasis on plant protection. By contrast, root systems with few lateral and actively growing meristems are relatively invulnerable to pathogen attack, but these root systems are poor foragers for phosphorus. Here, AM symbioses are likely to have evolved with an emphasis on phosphorus capture. Of course, even this more sophisticated view of AM function is unlikely to be the whole story: other aspects of AM ecology, such as protection from herbivores and toxic metals, may well vary in ways unrelated to root architecture.

13.8.3 Ericoid mycorrhizas

Heathlands exist in environments characterized by soils with low levels of available plant nutrients, often as a result of regular fires in which, for example, up to 80% of the nitrogen that has accumulated between fires may be lost. It is unsurprising, therefore, that heathlands are dominated by many plants that have evolved an association with ericoid mycorrhizal fungi (Read, 1996). This enables them to facilitate the extraction of nitrogen and phosphorus from the superficial layers of detrital material generated by the plants. Indeed, the conservation of natural heathlands is threatened by nitrogen supplementation and fire control, which allow colonization and domination by grasses that would otherwise be unable to exist in these impoverished environments.

The ericoid mycorrhizal root itself is anatomically simple compared to other mycorrhizas, characterized by a reduction of its vascular and cortical tissues, by the absence of root hairs, and by the presence of swollen epidermal cells occupied by mycorrhizal fungi. As a result, the individual roots are delicate structures, often referred to as 'hair-roots'; collectively the hair-roots form a dense fibrous root system, the bulk of which is concentrated towards the surface of the soil profile (Pate, 1994). The fungi are effective, unlike the plants alone, in absorbing nitrate, ammonium and phosphate ions that have been mobilized by other decomposers in the soil (see Chapter 11), but crucially they are also 'saprotrophic'. They are therefore able to compete directly with the other decomposers in liberating nitrogen and phosphorus from the organic residues in which most of these elements are locked up in heathland ecosystems (Read, 1996). A mutualism can thus be seen, again, to be woven into a larger web of interactions: the symbiont enhances its contribution to the host by making a preemptive competitive strike for scarce inorganic resources, and its own competitive ability is presumably enhanced in turn by the physiological support provided by its host.

13.9 Fungi with algae: the lichens

Of the 70,000 or so species of fungus that are known, approximately 20% are 'lichenized' (Palmqvist, 2000). Lichens are nutritionally specialized fungi (the so-called 'mycobiont' component) that have escaped from their normal way of life into a mutualistic association with a 'photobiont'. In around 90% of lichen species the photobiont is an alga, which provides carbon compounds to the mycobiont through photosynthesis. In some cases, the photobiont is a cyanobacterium, which may also provide fixed nitrogen to the association. In a relatively few, 'tripartite' lichen species (around 500) both an alga and a cyanobacterium are involved. Lichenized fungi belong to diverse taxonomic groups and the mutualistic algae to 27 different genera. Presumably, the lichen habit has evolved many times.

mycobionts and
photobionts

The photobionts are located extracellularly between the fungal hyphae, in a thin layer near the upper surface. Together, the two components form an integrated 'thallus' but the photobiont makes up only about 3–10% by weight. The advantage to the photobiont in the association, if any, has not been established clearly. All lichenized algal species, for example, can also occur free-living outside their association with their mycobiont. It may be that they are 'captured' by the fungus and exploited without any recompense. However, some of the species (e.g. of algal genus *Trebouxia*) are rare in their free-living form but very common in lichens, suggesting that there is something special about life in their mycobiont that they need. Moreover, since minerals, including nitrogen, are largely 'captured' from what is deposited



Figure 13.18 A variety of lichen species on a tree trunk. Reproduced by permission of Vaughan Fleming/Science Photo Library.

directly onto the lichen, often in rainwater and from the flow and drip down the branches of trees, and since the surface and biomass are largely fungal, the mycobiont must contribute the vast bulk of these minerals.

parallels with higher plants

Hence, the mutualistic pairs (and threesomes) in lichens provide two striking parallels with higher plants.

There is a structural parallel: in plants, the photosynthetic chloroplasts (see also Section 13.12) are similarly concentrated close to light-facing surfaces. There is also a functional parallel. The economy of a plant relies on carbon produced largely in the leaves and nitrogen absorbed principally through the roots, with a relative shortage of carbon giving rise to shoot growth at the expense of roots, and a shortage of nitrogen leading to root growth at the expense of shoots. Likewise, in lichens, the synthesis of carbon-fixing photobiont cells is inhibited by a relative shortage of nitrogen in the mycobiont,

but synthesis is stimulated when the carbon supply is limiting (Palmqvist, 2000).

Lichenization, then, gives the mycobiont and the photobiont, between them, the functional role of higher plants, but in so doing it extends the ecological range of both partners onto substrata (rock surfaces, tree trunks) and into regions (arid, arctic and alpine) that are largely barred to higher plants. Indeed, it has been claimed that lichens dominate 8% of terrestrial communities, both in terms of abundance and species diversity. However, all lichens grow slowly: the colonizers of rock surfaces rarely extend faster than 1–5 mm year⁻¹. They are, though, very efficient accumulators of the mineral cations that fall or drip onto them, and this makes them particularly sensitive to environmental contamination by heavy metals and fluoride. Hence, they are amongst the most sensitive indicators of environmental pollution. The ‘quality’ of an environment in humid regions can be judged rather accurately from the presence or absence of lichen growth on tombstones and tree trunks.

One remarkable feature in the life of the lichenized fungi is that the growth form of the fungus is usually profoundly changed when the alga is present. When the fungi are cultured in isolation from the algae, they grow slowly in compact colonies, much like related free-living fungi; but in the presence of the algal symbionts they take on a variety of morphologies (Figure 13.18) that are characteristic of specific algal–fungal partnerships. In fact, the algae stimulate morphological responses in the fungi that are so precise that the lichens are classified as distinct species, and a cyanobacterium and an alga, for example, may elicit quite different morphologies from the same fungus.

remarkable morphological responses on the fungi

13.10 Fixation of atmospheric nitrogen in mutualistic plants

The inability of most plants and animals to fix atmospheric nitrogen is one of the great puzzles in the process of evolution, since nitrogen is in limiting supply in many habitats. However, the ability to fix nitrogen is widely though irregularly distributed amongst both the eubacteria (‘true’ bacteria) and the archaea (archaeobacteria), and many of these have been caught up in tight mutualisms with systematically quite different groups of eukaryotes. Presumably such symbioses have evolved a number of times independently. They are of enormous ecological importance because of nitrogen’s frequent importance (Sprent & Sprent, 1990).

The nitrogen-fixing bacteria that have been found in symbioses (not necessarily mutualistic) are members of the following taxa.

the range of nitrogen-fixing bacteria

- 1 Rhizobia, which fix nitrogen in the root nodules of most leguminous plants and just one nonlegume, *Parasponia* (a member of the family Ulmaceae, the elms). At least three genera are recognized: *Rhizobium*, *Bradyrhizobium* and *Azorhizobium*, which are so distinct that they should perhaps be in different families (Sprent & Sprent, 1990), and between them they may comprise 10^4 or more species.
- 2 Actinomycetes of the genus *Frankia*, which fix nitrogen in the nodules (actinorhiza) of a number of nonleguminous and mainly woody plants, such as alder (*Alnus*) and sweet gale (*Myrica*).
- 3 Azotobacteriaceae, which can fix nitrogen aerobically and are commonly found on leaf and root surfaces.
- 4 Bacillaceae, such as *Clostridium* spp., which occur in ruminant feces, and *Desulfotomaculum* spp., which fix nitrogen in mammalian guts.
- 5 Enterobacteriaceae, such as *Enterobacter* and *Citrobacter*, which occur regularly in intestinal floras (e.g. of termites) and occasionally on leaf surfaces and on root nodules.
- 6 Spirillaceae, such as *Spirillum lipiferum*, which is an obligate aerobe found on grass roots.
- 7 Cyanobacteria of the family Nostocaceae, which are found in association with a remarkable range (though rather few species) of flowering and nonflowering plants (see Section 13.10.3), and which we recently met as photobionts in lichens.

Of these, the association of the rhizobia with legumes is the most thoroughly studied, because of the huge agricultural importance of legume crops.

13.10.1 Mutualisms of rhizobia and leguminous plants

several steps to a liaison

The establishment of a liaison between rhizobia and legume plants proceeds by a series of reciprocating steps. The bacteria occur in a free-living state in the soil and are stimulated to multiply by root exudates and cells that have been sloughed from roots as they develop. These exudates are also responsible for switching on a complex set of genes in the rhizobia (*nod* genes) that control the process that induces nodulation in the roots of the host. In a typical case, a bacterial colony develops on the root hair, which then begins to curl and is penetrated by the bacteria. The host responds by laying down a wall that encloses the bacteria and forms an ‘infection thread’, within which the rhizobia proliferate extracellularly. This grows within the host root cortex, and the host cells divide in advance of it, beginning to form a nodule. Rhizobia in the infection thread cannot fix nitrogen, but some are released into the host meristem cells. There, surrounded by a host-derived peribacteroid membrane, they differentiate into ‘bacteroids’ that can fix nitrogen. In some species, those with ‘indeterminate’ growth like the rhizobia of

the pea (*Pisum sativum*), the bacteroids themselves are unable to reproduce further. Only undifferentiated rhizobia are released back into the soil to associate with another root when the original root senesces. By contrast, in species with ‘determinate’ growth like those of the soybean (*Glycine max*), bacteroids survive root senescence and can then invade other roots (Kiers *et al.*, 2003).

A special vascular system develops in the host, supplying the products of photosynthesis to the nodule tissue and carrying away fixed nitrogen compounds (very often the amino acid asparagine) to other parts of the plant (Figure 13.19). The nitrogen-fixing nitrogenase enzyme accounts for up to 40% of the protein in the nodules and depends for its activity on a very low oxygen tension. A boundary layer of tightly packed cells within the nodule serves as a barrier to oxygen diffusion. A hemoglobin (leghemoglobin) is formed within the nodules, giving the active nodules a pink color. It has a high affinity for oxygen and allows the symbiotic bacteria to respire aerobically in the virtually anaerobic environment of the nodule. Indeed, wherever nitrogen-fixing symbioses occur, at least one of the partners has special structural (and usually also biochemical) properties that protect the anaerobic nitrogenase enzyme from oxygen, yet allow normal aerobic respiration to occur around it.

13.10.2 Costs and benefits of rhizobial mutualisms

The costs and benefits of this mutualism need to be considered carefully. From the plant’s point of view, we need to compare the energetic costs of alternative processes by which supplies of fixed nitrogen might be obtained. The route for most plants is direct from the soil as nitrate or ammonium ions. The metabolically cheapest route is the use of ammonium ions, but in most soils ammonium ions are rapidly converted to nitrates by microbial activity (nitrification). The energetic cost of reducing nitrate from the soil to ammonia is about 12 mol of adenosine triphosphate (ATP) per mol of ammonia formed. The mutualistic process (including the maintenance costs of the bacteroids) is energetically slightly more expensive to the plant: about 13.5 mol of ATP. However, to the costs of nitrogen fixation itself we must also add the costs of forming and maintaining the nodules, which may be about 12% of the plant’s total photosynthetic output. It is this that makes nitrogen fixation energetically inefficient. Energy, though, may be much more readily available for green plants than nitrogen. A rare and valuable commodity (fixed nitrogen) bought with a cheap currency (energy) may be no bad bargain. On the other hand, when a nodulated legume is provided with nitrates (i.e. when nitrate is not a rare commodity) nitrogen fixation declines rapidly.

The benefits to the rhizobia are more problematic from an evolutionary point of view, especially for those with indeterminate growth, where the rhizobia that have become bacteroids can

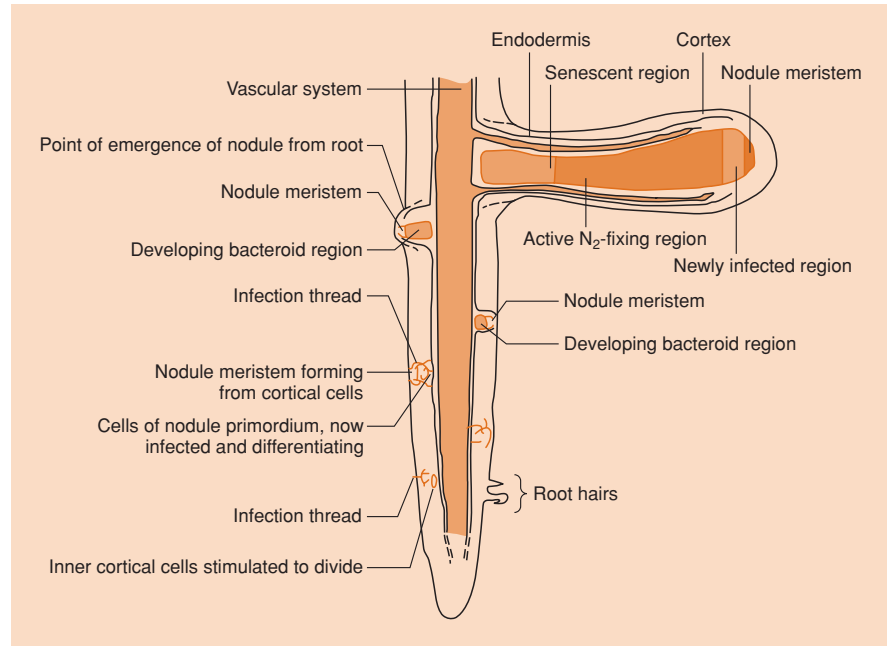


Figure 13.19 The development of the root nodule structure during the course of development of infection of a legume root by *Rhizobium*. (After Sprent, 1979.)

fix nitrogen but cannot reproduce. Hence, they cannot themselves benefit from the symbiosis, since ‘benefit’ must express itself, ultimately, as an increased reproductive rate (fitness). The rhizobia in the infection thread are capable of reproduction (and are therefore able to benefit), but they cannot fix nitrogen and are therefore not themselves involved in a mutualistic interaction. However, since the rhizobia are clonal, the bacteroids and the cells in the infection thread are all part of the same, single genetic entity. The bacteroids, therefore, by supporting the plant and generating a flow of photosynthates, can benefit the cells of the infection thread, and hence benefit the clone as a whole, in much the same way as the cells in a bird’s wing can bring benefit, ultimately, to the cells that produce its eggs – and hence to the bird as a whole.

why no cheating?

One puzzle, though, since the rhizobia associated with a particular plant are typically a mixture of clones, is why individual clones do not ‘cheat’: that is, derive benefits from the plant, which itself derives benefit from the rhizobia in general, without themselves entering fully into the costly enterprise of fixing nitrogen. Indeed, we can see that this question of *cheating* applies to many mutualisms, once we recognize that they are, in essence, cases of mutual exploitation. There would be evolutionary advantage in exploiting without being exploited. Perhaps the most obvious answer is for the plant (in this case) to monitor the performance of the rhizobia and apply ‘sanctions’ if they cheat. This, clearly, will provide evolutionary stability to the mutualism by preventing cheats from escaping the interaction, and evidence for such sanctioning has indeed been found for a legume–rhizobium mutualism (Kiers *et al.*, 2003). A normally mutualistic rhizobium strain was prevented from cooperating (fixing nitrogen) by

growing its soybean host in an atmosphere in which air (80% nitrogen, 20% oxygen) was replaced with approximately 80% argon, 20% oxygen and only around 0.03% nitrogen, reducing the rate of nitrogen fixation to around 1% of normal levels. Thus, the rhizobium strain was forced to cheat. In experiments at the whole plant, the part-root and the individual nodule level, the reproductive success of the noncooperating rhizobia was decreased by around 50% (Figure 13.20). Noninvasive monitoring of the plants indicated that they were applying sanctions by withholding oxygen from the rhizobia. Cheating did not pay.

13.10.3 Nitrogen-fixing mutualisms in nonleguminous plants

The distribution of nitrogen-fixing symbionts in nonleguminous higher plants is patchy. A genus of actinomycete, *Frankia*, forms symbioses (actinorrhiza) with members of at least eight families of flowering plants, almost all of which are shrubs or trees. The nodules are usually hard and woody. The best known hosts are the alder (*Alnus*), sea buckthorn (*Hippophaë*), sweet gale (*Myrica*), she-oak (*Casuarina*) and the arctic/alpine shrubs *Arctostaphylos* and *Dryas*. *Ceanothus*, which forms extensive stands in Californian chaparral, also develops *Frankia* nodules. Unlike rhizobia, the species of *Frankia* are filamentous and produce specialized vesicles and sporangia that release spores. Whilst the rhizobia rely on their host plant to protect their nitrogenase from oxygen, *Frankia* provides its own protection in the walls of the vesicles, which are massively thickened with as many as 50 monolayers of lipids.

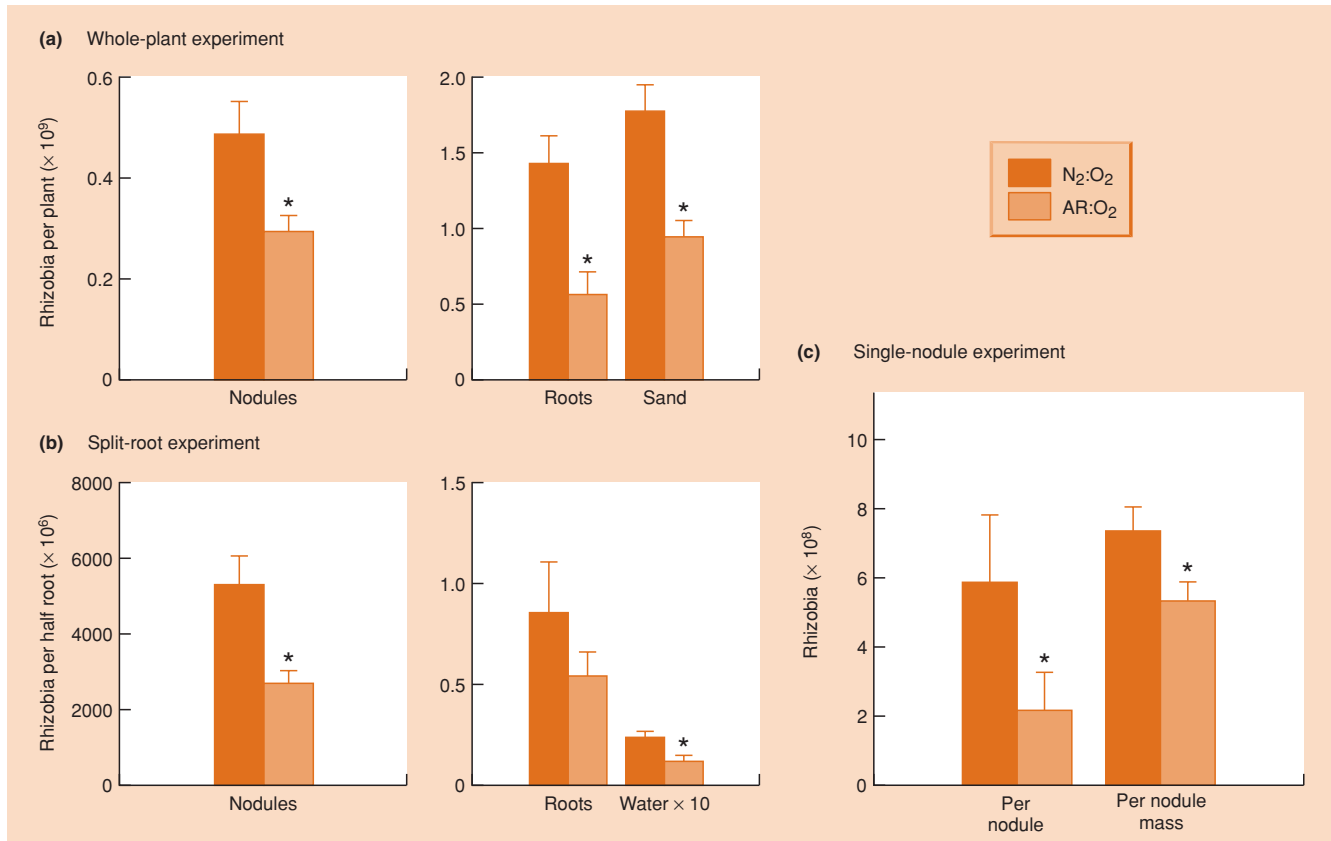


Figure 13.20 The number of rhizobia grew to much larger numbers when allowed to fix nitrogen in normal air (N₂ : O₂) than when prevented from doing so by manipulation of the atmosphere (Ar : O₂). (a) When the different treatments were applied at the whole plant level, there were greater numbers within the nodules (left; $P < 0.005$) and on the root surface (right; both $P < 0.01$) and in the surrounding sand ($P < 0.01$). $n = 11$ pairs; bars are standard errors. (b) When the different treatments were applied to different parts of the same root system, there were greater numbers within the nodules (left; $P < 0.001$) and for those in the surrounding water (right; $P < 0.01$), but not significantly so for those on the root surface. $n = 12$ plants; bars are standard errors. (c) When the different treatments were applied to individual nodules from the same root system, there were greater numbers on a per nodule basis ($P < 0.05$) and a per nodule mass basis ($P < 0.01$). $n = 6$ experiments; bars are standard errors. (After Keirs *et al.*, 2003.)

Cyanobacteria form symbioses with three genera of liverwort (*Anthoceros*, *Blasia* and *Clavicularia*), with one fern (the free-floating aquatic *Azolla*), with many cycads (e.g. *Encephalartos*) and with all 40 species of the flowering plant genus *Gunnera*, but with no other flowering plants. In the liverworts, the cyanobacteria *Nostoc* live in mucilaginous cavities and the plant reacts to their presence by developing fine filaments that maximize contact with it. *Nostoc* is found at the base of the leaves of *Gunnera*, in the lateral roots of many cycads, and in pouches in the leaves of *Azolla*.

13.10.4 Interspecific competition

The mutualisms of rhizobia and legumes (and other nitrogen-fixing mutualisms) must not be seen as isolated interactions between

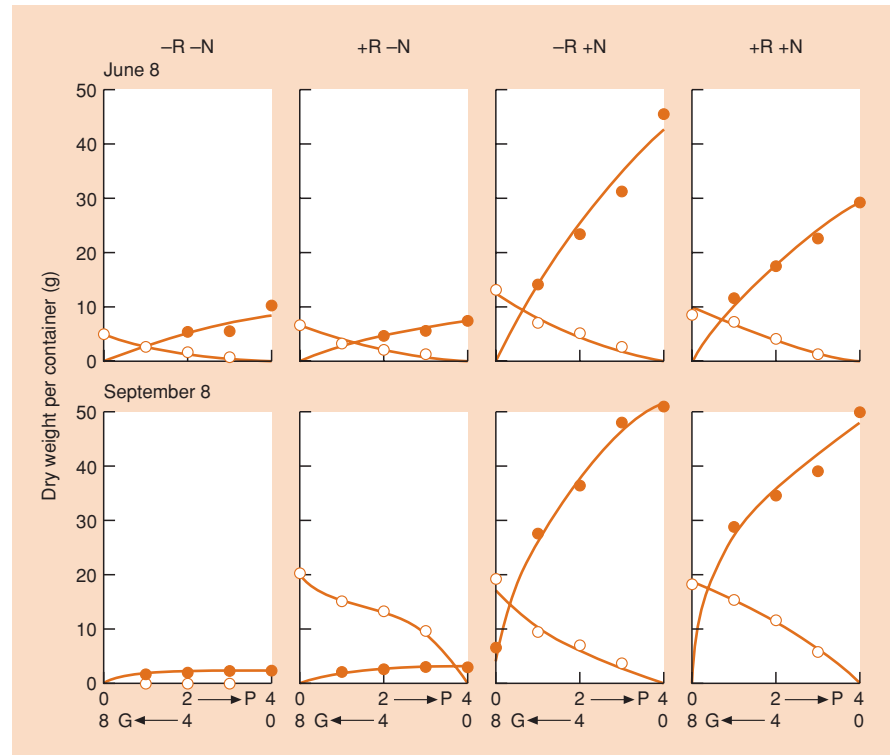
bacteria and their own host plants. In nature, legumes normally form mixed stands in association with nonlegumes. These are potential competitors with the legumes for fixed nitrogen (nitrates or ammonium ions in the soil). The nodulated legume sidesteps this competition by its access to a unique source of nitrogen. It is in this ecological context that nitrogen-fixing mutualisms gain their main advantage. Where nitrogen is plentiful, however, the energetic costs of nitrogen fixation often put the plants at a competitive disadvantage.

Figure 13.21, for example, shows the results of a classic experiment in which soybeans (*Glycine soja*, a legume) were grown in mixtures with *Paspalum*,

a classic 'replacement series'

a grass. The mixtures either received mineral nitrogen, or were inoculated with *Rhizobium*, or received both. The experiment

Figure 13.21 The growth of soybeans (*Glycine soja*, G, ○) and a grass (*Paspalum*, P, ●) grown alone and in mixtures with and without nitrogen fertilizer and with and without inoculation with nitrogen-fixing *Rhizobium*. The plants were grown in pots containing 0–4 plants of the grass and 0–8 plants of *Glycine*. The horizontal scale on each figure shows the mass of plants of the two species in each container. –R –N, no *Rhizobium*, no fertilizer; +R –N, inoculated with *Rhizobium* but no fertilizer; –R +N, no *Rhizobium* but nitrate fertilizer was applied; +R +N, inoculated with *Rhizobium* and nitrate fertilizer was supplied. (After de Wit *et al.*, 1966.)



was designed as a ‘replacement series’ (see Section 8.7.2), which allows us to compare the growth of pure populations of the grass and legume with their performances in the presence of each other. In the pure stands of soybean, yield was increased very substantially either by inoculation with *Rhizobium* or by application of fertilizer nitrogen, or by receiving both. The legumes can use either source of nitrogen as a substitute for the other. The grass, however, responded only to the fertilizer. Hence, when the species competed in the presence of *Rhizobium* alone, the legume contributed far more to the overall yield than did the grass: over a succession of generations, the legume would have outcompeted the grass. When they competed in soils supplemented with fertilizer nitrogen, however, whether or not *Rhizobium* was also present, it was the grass that made the major contribution: long term, it would have outcompeted the legume.

Quite clearly, then, it is in environments deficient in nitrogen that nodulated legumes have a great advantage over other species. But their activity raises the level of fixed nitrogen in the environment. After death, legumes augment the level of soil nitrogen on a very local scale with a 6–12-month delay as they decompose. Thus, their advantage is lost – they have improved the environment of their competitors, and the growth of associated grasses will be favored in these local patches. Hence, organisms that can fix atmospheric nitrogen can be thought of as locally suicidal. This is one reason why it is very difficult to grow

repeated crops of pure legumes in agricultural practice without aggressive grass weeds invading the nitrogen-enriched environment. It may also explain why leguminous herbs or trees usually fail to form dominant stands in nature.

Grazing animals, on the other hand, continually remove grass foliage, and the nitrogen status of a grass patch may again decline to a level at which the legume may once more be at a competitive advantage. In a stoloniferous legume, such as white clover, the plant is continually ‘wandering’ through the sward, leaving behind it local grass-dominated patches, whilst invading and enriching with nitrogen new patches where the nitrogen status has become low. The symbiotic legume in such a community not only drives its nitrogen economy but also some of the cycles that occur within its patchwork (Cain *et al.*, 1995).

13.10.5 Nitrogen-fixing plants and succession

An ecological succession (treated in much more detail in Chapter 17) is the directional replacement of species by other species at a site. A shortage of fixed nitrogen commonly hinders the earliest stages of the colonization of land by vegetation: the initial stages of a succession on open land. Some fixed nitrogen will be contributed in rain after thunderstorms, and some may be blown in from other more established areas, but nitrogen-fixing

organisms such as bacteria, cyanobacteria and lichens are important pioneer colonizers. Higher plants with nitrogen-fixing symbionts, however, are rarely pioneers. The reason appears to be that open land is usually colonized first by plants with light, dispersible seeds. A legume seedling, however, depends on fixed nitrogen in its seed reserves and the soil before it can grow to a stage where it can nodulate and fix nitrogen for itself. It is likely, therefore, that only large-seeded legumes carry enough fixed nitrogen to carry them through the establishment phase, and species with such large seeds will not have the dispersibility needed to be pioneers (Grubb, 1986; see also Sprent & Sprent, 1990).

Finally, note that since symbiotic nitrogen fixation is energetically demanding, it is not surprising that most of the higher plant species that support nitrogen-fixing mutualists are intolerant of the shade that is characteristic of the late stages of successions. Higher plants with nitrogen-fixing mutualists are seldom in at the beginning of a succession and they seldom persist to the end.

13.11 Models of mutualisms

Several of the previous chapters on interactions have included a section on mathematical models. This is perhaps a good time to remind ourselves why this was – because the models, by separating essence from detail, were able to provide insights that would not be apparent from a catalog of actual examples. For modeling to be a success, then, it is imperative that the ‘essence’ is correctly identified. What is the essence of a mutualism? One might imagine it to be that each partner has a positive influence on the fitness of the other partner. At first sight, therefore, we might imagine that an appropriate model for a mutualistic interaction would simply replace the negative contributions in models of two-species competition (see Chapter 8) with positive contributions. However, such a model leads to absurd solutions in which both populations explode to unlimited size (May, 1981) because it places no limits on the carrying capacity of either species, which would therefore increase indefinitely. In practice, intraspecific competition for limiting resources must eventually determine a maximum carrying capacity for any mutualist population, even if the population of the partner mutualist is present in excess (Dean, 1983). Thus, a plant whose growth is limited by a shortage of fixed nitrogen may be released into faster growth by mutualism with a nitrogen-fixing partner, but its faster growth must soon become constrained by a shortage of some other limiting resource (e.g. water, phosphate, radiant energy).

This returns us to points made at the start of the chapter: that the essence of mutualism is subtler than ‘mutual benefit’. Rather, instead of thinking of each partner as a benefit to the other, without qualification, it is better to think of each partner exploiting the other, with benefits to be gained but also costs to pay.

And to recognize, too, that the balance of benefits and costs can alter – with changing conditions, changing resource levels, the abundance of either partner, and the presence or abundance of other species. Even the simplest of models, therefore, would have to have terms not signifying a ‘positive contribution’, but terms that could be positive or negative according to the state of some other part of the model community: not simple at all compared to the models described, and found useful, in previous chapters.

In a sense, then, turning to models does prove helpful at this point. Models of predator–prey and competitor pairs in isolation capture an essence of predator–prey and competitive interactions. The fact that models of isolated pairs of ‘mutualists’ cannot do so re-emphasizes that mutualism is, in terms of population dynamics, *essentially* an interaction that should only be viewed within the broader context of a larger community. We have seen this earlier in the chapter, for example in ants and aphids in the presence and absence of aphid predators, in *Symbiodinium* taxa coexisting in coral, and in the legume–*Rhizobium* mutualism, which brings its great advantage to the legume when it is competing with some other plant (e.g. a grass) for limited nitrogen from the soil.

This point has been captured in a mathematical model, in which a two-species bee–plant pollination mutualism (Figure 13.22a) was examined embedded in a community that also contained another species of plant and a species of bird that preyed on the bees (Figure 13.22b) (Ringel *et al.*, 1996). The bees could either take nectar and pollen from the plants but fail to pollinate them (predator–prey) or could pollinate them successfully (mutualistic). The model of the simple mutualistic pair (Figure 13.22a) was, as described above, intrinsically unstable. The pair could only persist if the strength of intraspecific competition exceeded that of mutualism: the more mutualistic the interaction became, the more unstable it was. Taken at face value, such results seem to suggest that mutualisms will be rare (though we have seen they are not), since the conditions for their existence are restricted.

A quite different picture emerges, however, once the pair is embedded in the larger assemblage (Figure 13.22b). Using a variety of measures, one of which is illustrated and explained in Figure 13.22c, it was apparent that mutualisms tended to increase the chances of the assemblage’s persistence. Clearly, there is no necessary paradox between the widespread occurrence of mutualistic interactions in nature and their effects in model assemblages of species. Equally clearly, though, while model assemblages are inevitably simple (e.g. just five species), the effects of mutualistic interactions in nature can easily be misjudged if they are *too* simple (i.e. the mutualistic pair alone).

models of two-species mutualisms stress the importance of knowing the broader context

a bird, a bee and two plants

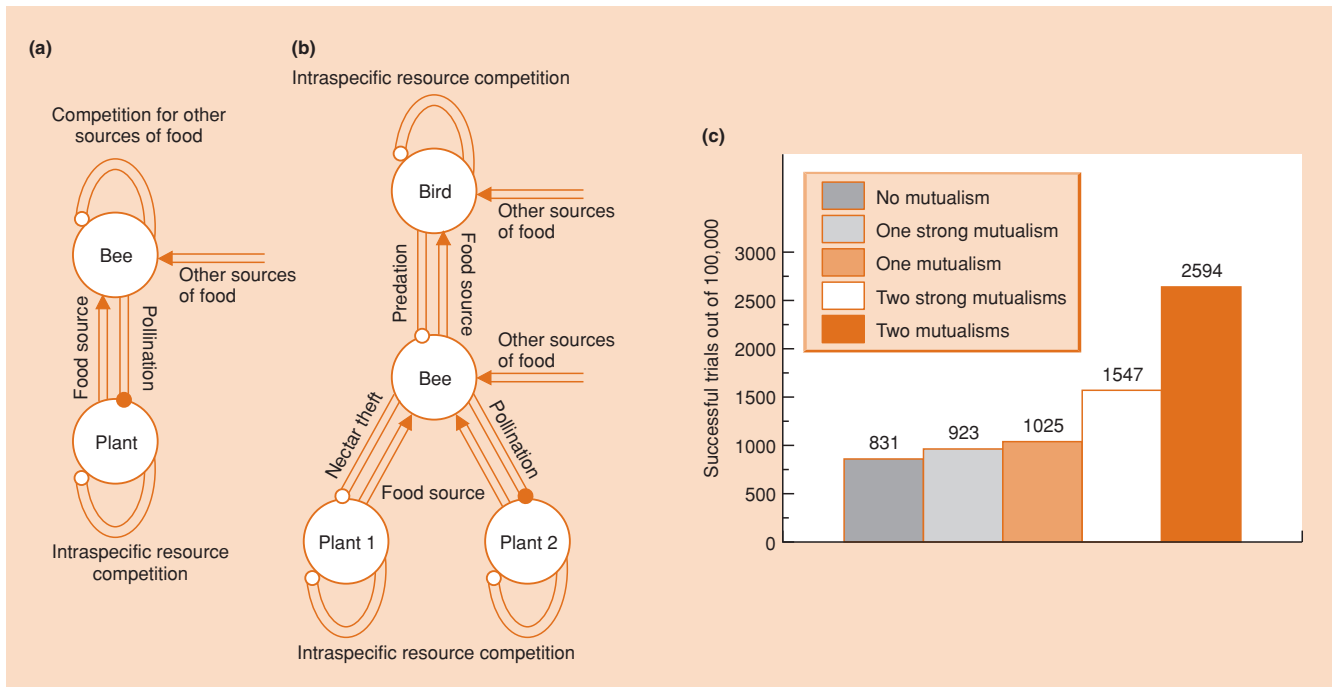


Figure 13.22 (a) A model two-species bee–plant mutualism. Both species are also subject to intraspecific competition. Filled arrowheads indicate a positive interaction, either a resource → consumer (pointed) or a pollination (round) interaction; open arrowheads indicate negative interactions, either consumer → resource or intraspecific competition. (b) The bee and plant embedded in a community with another plant and a bird predator of the bees. The plants suffer intraspecific competition but do not compete with one another. The birds suffer intraspecific competition but the bees do not. The bees take pollen and nectar from both plant species and either fail to pollinate them (predator–prey) or pollinate them successfully (mutualistic). In the figure, the interaction with plant 1 is predator–prey and that with plant 2 is mutualistic, but cases were examined in which neither, one or both were mutualistic. (c) Comparison of persistence in the possible assemblages in (b). Persistence of an assemblage is the maintenance of all species at positive population densities. The bars indicate the number that persisted when the dynamics of each assemblage were simulated 10,000 times, with the strengths of each interaction given by values generated randomly within defined bounds. In a ‘strong mutualism’, the strength of interaction could be up to twice that in a ‘mutualism’. Mutualisms greatly increased the chances of persistence; two-tailed *t*-test of persistence versus no mutualism: one mutualism ($t = 4.52$, $P < 0.001$), one strong mutualism ($t = 2.21$, $P < 0.05$), two mutualisms ($t = 30.46$, $P < 0.001$), two strong mutualisms ($t = 14.78$, $P < 0.001$). (After Ringel *et al.*, 1996.)

13.12 Evolution of subcellular structures from symbioses

We have seen in this chapter that there is remarkable variety in the types of association that may be regarded as symbiotic – many of them shown clearly to be mutualistic. They extend from patterns of behavior linking two very different organisms that spend parts of their lives apart, through the microbial communities of the vertebrate gut (strictly external to the body tissues), to the intercellular ectomycorrhizas and lichens, and the intracellular dinoflagellate algae of corals and mycocyte bacteria of insects. We end this chapter by examining how an ecological interaction – mutualism – may lie at the heart of biological patterns operating on the longest evolutionary timescales.

It is now generally accepted that the origin of the various sorts of eukaryotes from more primitive ancestors has progressed at least in part through the

the serial endosymbiosis theory

inextricable merging of partners in a symbiosis. This view was championed especially by Margulis (1975, 1996) in the ‘serial endosymbiosis theory’ (Figure 13.23a). The aim is to understand the relationships between the three ‘domains’ of living organism: the archaeobacteria or Archaea (many of them now ‘extremophiles’, living at high temperatures, low pHs and so on), the ‘true’ bacteria (Eubacteria) and the eukaryotes (Katz, 1998). One suggested first step (estimated to have occurred around 2 billion years ago) was the merger of archaeal and bacterial (spirochete) cells in an anaerobic symbiosis. The former brought its nucleocytoplasm and the latter brought its swimming motility, thus explaining the

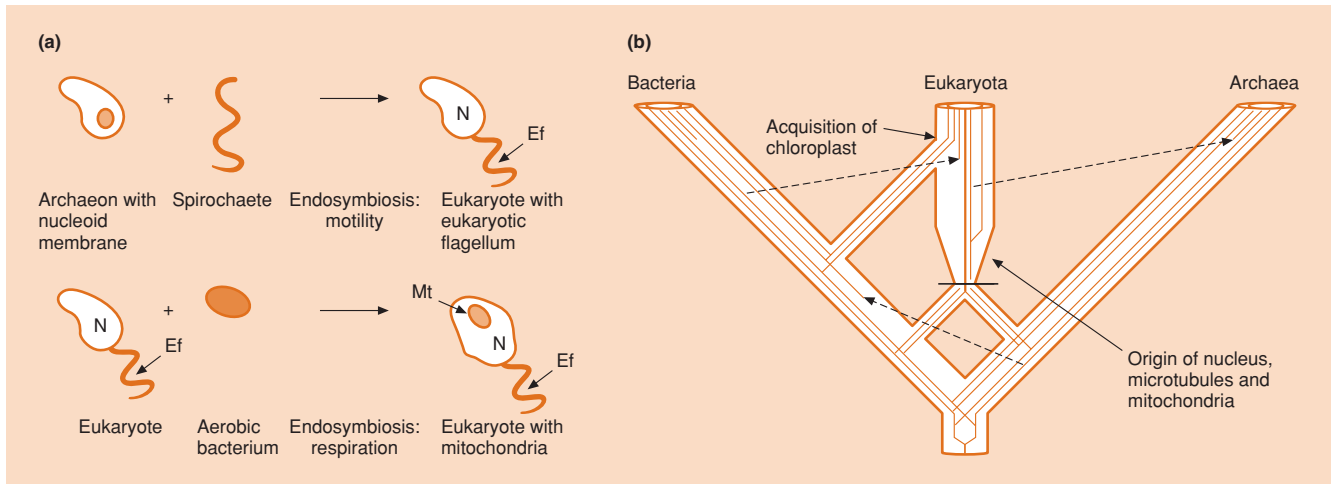


Figure 13.23 (a) The first two steps in the serial endosymbiosis theory for the origin of the eukaryotic cell. Ef, eukaryote flagellum; Mt, mitochondrion; N, nucleus. (b) A model for the origin of eukaryotes indicating a symbiosis between archaeal and bacterial lineages, and the possible simultaneous origin of nuclei, microtubules and mitochondria in eukaryotes. Bold lines represent lineage boundaries; pale lines are gene genealogies; broken arrows are possible lateral transfers of individual genes. (After Katz, 1998.)

chimeric nature – the mix of archaeal and bacterial features – of the proteins and the genetic material of even the most primitive eukaryotes. Subsequently, some of these chimeras incorporated aerobic bacteria that were the forerunners of mitochondria, to become aerobic eukaryotes from which all other eukaryotes have evolved. Some of these later acquired phototrophic cyanobacteria that were the forerunners of chloroplasts, providing the stock from which the algae and higher plants evolved.

In fact, the serial endosymbiosis theory is merely one of several seeking to link the three domains and recreate the origins of the eukaryotes (Katz, 1998). A suggestion, for instance, that the most primitive eukaryotes have lost mitochondria, rather than never having had them, calls into question the whole sequential nature of eukaryote origins. It may also be that the ‘lateral transfer’ of individual genes (from one evolutionary lineage to another) has been more pervasive over evolutionary time than was previously imagined, so that the branching tree of life is in fact much more of a tangled web (Figure 13.23b). No doubt, as further evidence accumulates, these competing theories will themselves evolve further, both through progression and the lateral transfer of ideas. What they share, however, is the idea that mutualistic symbioses, beyond their ecological importance, lie at the heart of some of the most fundamental steps in evolution.

Summary

We start by distinguishing mutualism, symbiosis and commensalism and emphasizing that mutualism is best seen as reciprocal exploitation not a cosy partnership.

Mutualisms are examined in a progression: from those where the association is behavioral, through intimate symbioses in which one partner enters between or within another’s cells, to those where organelles have entered into such intimate symbioses within the cells of their hosts that they cannot be regarded as distinct organisms.

‘Cleaner’ fish feed on ectoparasites, bacteria and necrotic tissue from the body surface of ‘client’ fish. The cleaners gain a food source and the clients are protected from infection. Many ant species protect plants from predators and competitors, while themselves feeding from specialized parts of the plants, though careful experiments are necessary to show that the plants themselves benefit.

Many species, including humans, culture crops or livestock from which they feed. Ants farm many species of aphids in return for sugar-rich secretions, though experiments demonstrate that there can be both costs and benefits for the aphids. Many ants and beetles farm fungi that give them access to otherwise indigestible plant material, and in some cases a three-way mutualism is established with actinomycetes that protect the fungi from pathogens.

Very many plant species use animals to disperse their seeds and pollen. We emphasize the importance of insect pollinators and the coevolutionary pressures generating a range from generalists to ultraspecialists. We also discuss brood site pollinations, of figs and yuccas, by fig wasps and yucca moths that rear their larvae in the fruits of the pollinated plant.

Many animals support a mutualistic microbiota within their guts, especially important in the digestion of cellulose. We outline the range of active sites, and the complex community of mutu-

alists, within the guts of a variety of vertebrates and of termites, focusing especially on the ruminants and noting the importance in many cases of refecation. We also describe insect mycetocyte symbioses, especially those between aphids and *Buchnera* species, through which microorganisms, mostly bacteria, living in specialized cells bring nutritional benefits to their insect hosts.

A number of aquatic invertebrates enter into mutualistic associations with photosynthetic algae, perhaps the most important of which are the reef-building corals. We focus especially on 'coral bleaching' – the whitening of corals as a result of the loss of the endosymbionts – and its possible relationship with global warming, and we emphasize the multi- (not two) species nature of these and many other mutualisms.

A wide variety of symbiotic associations are formed between higher plants and fungi. We concentrate on the mycorrhizas – intimate mutualisms between fungi and root tissues – possessed by most plants. We describe ectomycorrhizas, arbuscular mycorrhizas and ericoid mycorrhizas, noting the range of benefits that they can confer.

The biology is outlined of lichens, discussing the intimate associations between mycobiont fungi and phytobionts, mostly algae. Parallels with higher plants are particularly emphasized.

Mutualisms between plants and nitrogen-fixing bacteria are of enormous importance. We outline the range of these bacteria but focus mainly on the mutualisms of rhizobia and leguminous plants, describing the steps involved in establishing the liaison, the costs and benefits to both parties, and the role of the mutualism in determining the outcome of competition between legumes and other plants. This leads to a discussion of the part played by nitrogen-fixing plants in ecological successions.

We examine briefly some mathematical models of mutualisms, which re-emphasize the importance of looking beyond two focal species to the broader context.

Finally, we discuss the possibility that the origin of the various sorts of eukaryotes from more primitive ancestors has progressed at least in part through the inextricable merging of partners in mutualistic symbioses.



Chapter 14

Abundance

14.1 Introduction

Why are some species rare and others common? Why does a species occur at low population densities in some places and at high densities in others? What factors cause fluctuations in a species' abundance? These are crucial questions. To provide complete answers for even a single species in a single location, we might need, ideally, a knowledge of physicochemical conditions, the level of resources available, the organism's life cycle and the influence of competitors, predators, parasites, etc., as well as an understanding of how all these things influence abundance through their effects on the rates of birth, death and movement. In previous chapters, we have examined each of these topics separately. We now bring them together to see how we might discover which factors actually matter in particular examples.

counting is not enough

The raw material for the study of abundance is usually some estimate of population size. In its crudest form, this consists of a simple count. But this can hide vital information. As an example, picture three human populations containing identical numbers of individuals. One of these is an old people's residential area, the second is a population of young children, and the third is a population of mixed age and sex. No amount of attempted correlation with factors outside the population would reveal that the first was doomed to extinction (unless maintained by immigration), the second would grow fast but only after a delay, and the third would continue to grow steadily. More detailed studies, therefore, involve recognizing individuals of different age, sex, size and dominance and even distinguishing genetic variants.

estimates are usually deficient

Ecologists usually have to deal with estimates of abundance that are deficient. First, data may be misleading unless sampling is adequate over both space and time, and adequacy of either usually requires great

commitment of time and money. The lifetime of investigators, the hurry to produce publishable work and the short tenure of most research programs all deter individuals from even starting to conduct studies over extended periods of time. Moreover, as knowledge about populations grows, so the number of attributes of interest grows and changes; every study risks being out of date almost as soon as it begins. In particular, it is usually a technically formidable task to follow individuals in a population throughout their lives. Often, a crucial stage in the life cycle is hidden from view – baby rabbits within their warrens or seeds in the soil. It is possible to mark birds with numbered leg rings, roving carnivores with radiotransmitters or seeds with radioactive isotopes, but the species and the numbers that can be studied in this way are severely limited.

A large part of population theory depends on the relatively few exceptions where logistical difficulties have been overcome (Taylor, 1987). In fact, most of the really long-term or geographically extensive studies of abundance have been made of organisms of economic importance such as fur-bearing animals, game birds and pests, or the furry and feathered favorites of amateur naturalists. Insofar as generalizations emerge, we should treat them with great caution.

studied species may not be typical

14.1.1 Correlation, causation and experimentation

Abundance data may be used to establish correlations with external factors (e.g. the weather) or correlations between features within the abundance data themselves (e.g. correlating numbers present in the spring with those present in the fall). Correlations may be used to predict the future. For example, high intensities of the disease 'late blight' in the canopy of potato crops usually occur 15–22 days after a period in which the minimum

temperature is not less than 10°C and the relative humidity is more than 75% for two consecutive days. Such a correlation may alert the grower to the need for protective spraying.

Correlations may also be used to suggest, although not to prove, causal relationships. For example, a correlation may be demonstrated between the size of a population and its growth rate. The correlation may hint that it is the size of the population itself that causes the growth rate to change, but, ultimately, 'cause' requires a mechanism. It may be that when the population is high many individuals starve to death, or fail to reproduce, or become aggressive and drive out the weaker members.

density is an abstraction

In particular, as we have remarked previously, many of the studies that we discuss in this and other chapters have been concerned to detect 'density-dependent' processes, as if density itself is the cause of changes in birth rates and death rates in a population. But this will rarely (if ever) be the case: organisms do not detect and respond to the density of their populations. They usually respond to a shortage of resources caused by neighbors or to aggression. We may not be able to identify *which* individuals have been responsible for the harm done to others, but we need continually to remember that 'density' is often an abstraction that conceals what the world is like as experienced in the lives of real organisms.

Observing directly what is happening to the individuals may suggest more strongly still what causes a change in overall abundance. Incorporating observations on individuals into mathematical models of populations, and finding that the model population behaves like the real population, may also provide strong support for a particular hypothesis. But often, the acid test comes when it is possible to carry out a field experiment or manipulation. If we suspect that predators or competitors determine the size of a population, we can ask what happens if we remove them. If we suspect that a resource limits the size of a population, we can add more of it. Besides indicating the adequacy of our hypotheses, the results of such experiments may show that we ourselves have the power to determine a population's size: to reduce the density of a pest or weed, or to increase the density of an endangered species. Ecology becomes a predictive science when it can forecast the future: it becomes a management science when it can determine the future.

14.2 Fluctuation or stability?

Gilbert White's swifs

Perhaps the direct observations of abundance that span the greatest period of time are those of the swifs (*Micropus apus*) in the village of Selborne in southern England (Lawton & May, 1984). In one of the earliest published works on ecology, Gilbert White, who lived in the village, wrote of the swifs in 1778:

I am now confirmed in the opinion that we have every year the same number of pairs invariably; at least, the result of my inquiry has been exactly the same for a long time past. The number that I constantly find are eight pairs, about half of which reside in the church, and the rest in some of the lowest and meanest thatched cottages. Now, as these eight pairs – allowance being made for accidents – breed yearly eight pairs more, what becomes annually of this increase?

Lawton and May visited the village in 1983, and found major changes in the 200 years since White described it. It is unlikely that swifs had nested in the church tower for 50 years, and the thatched cottages had disappeared or had been covered with wire. Yet, the number of breeding pairs of swifs regularly to be found in the village was found to be 12. In view of the many changes that have taken place in the intervening centuries, this number is remarkably close to the eight pairs so consistently found by White.

Another example of a population showing relatively little change in adult numbers from year to year is seen in an 8-year study in Poland of the small, annual sand-dune plant *Androsace septentrionalis* (Figure 14.1a). Each year there was great flux within the population: between 150 and 1000 new seedlings per square meter appeared, but subsequent mortality reduced the population by between 30 and 70%. However, the population appears to be kept within bounds. At least 50 plants always survived to fruit and produce seeds for the next season.

The long-term study of nesting herons in the British Isles reported previously in Figure 10.23c reveals a picture of a bird population that has remained remarkably constant over long periods, but here, because repeated estimates were made, it is apparent that there were seasons of severe weather when the population declined precipitously before it subsequently recovered. By contrast, the mice in Figure 14.1b have extended periods of relatively low abundance interrupted by sporadic and dramatic interruptions.

14.2.1 Determination and regulation of abundance

Looking at these studies, and many others like them, some investigators have emphasized the apparent constancy of population sizes, while others have emphasized the fluctuations. Those who have emphasized constancy have argued that we need to look for stabilizing forces within populations to explain why they do not increase without bounds or decline to extinction. Those who have emphasized the fluctuations have looked to external factors, for example the weather, to explain the changes. Disagreements between the two camps dominated much of ecology in the middle third of the 20th century. By considering some of these

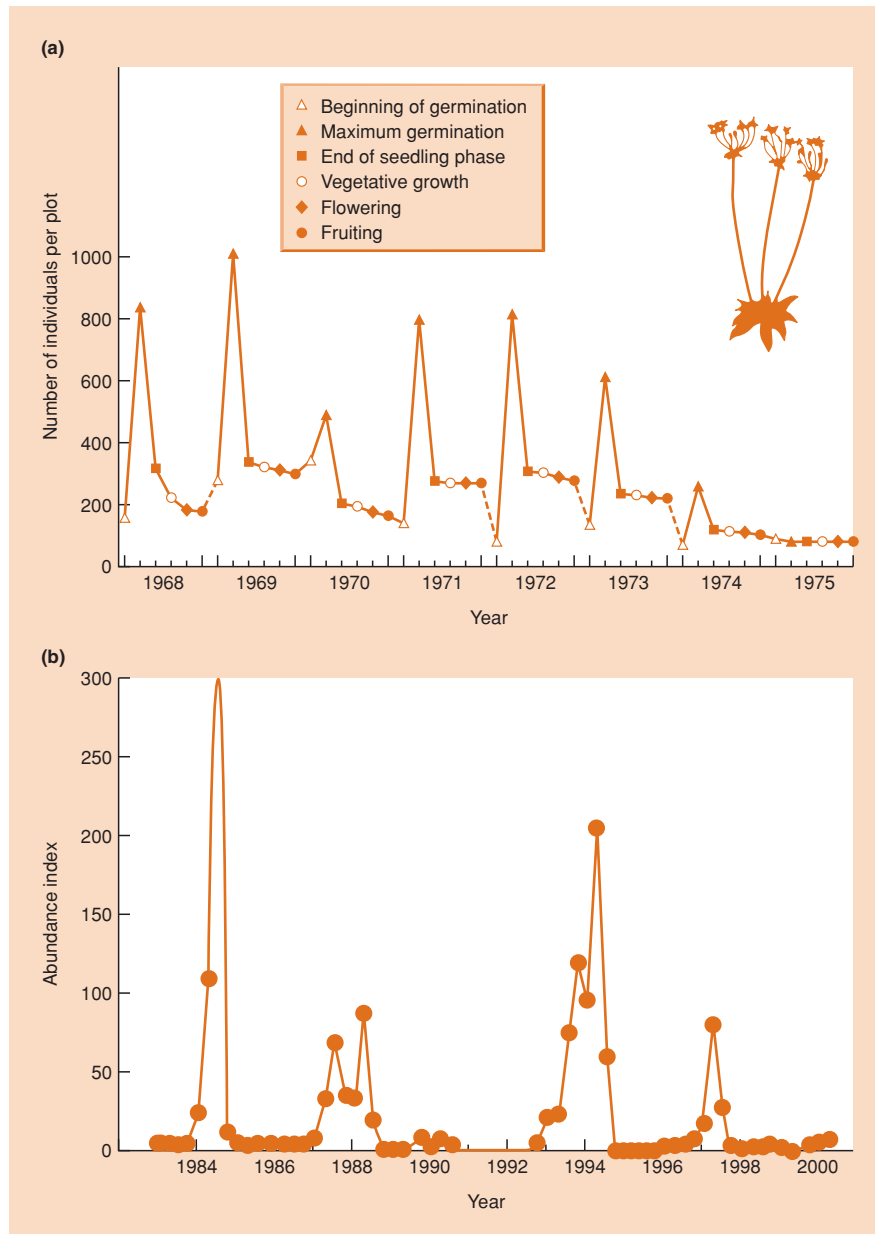


Figure 14.1 (a) The population dynamics of *Androsace septentrionalis* during an 8-year study. (After Symonides, 1979; a more detailed analysis of these data is given by Silvertown, 1982.) (b) Irregular irruptions in the abundance of house mice (*Mus domesticus*) in an agricultural habitat in Victoria, Australia, where the mice, when they irrupt, are serious pests. The 'abundance index' is the number caught per 100 trap-nights. In fall 1984 the index exceeded 300. (After Singleton *et al.*, 2001.)

arguments, it will be easier to appreciate the details of the modern consensus (see also Turchin, 2003).

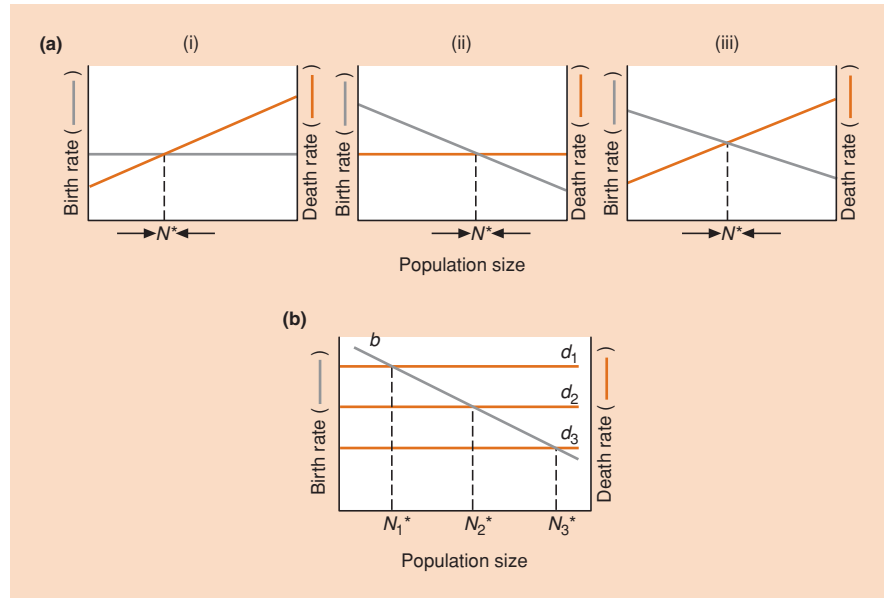
distinguishing the determination and regulation of abundance

First, however, it is important to understand clearly the difference between questions about the ways in which abundance is *determined* and questions about the way in which abundance is *regulated*. Regulation is the tendency of a population to decrease in size when it is above a particular level, but to increase in size when below that level. In other words, regulation of a population can, by definition, occur

only as a result of one or more density-dependent processes that act on rates of birth and/or death and/or movement. Various potentially density-dependent processes have been discussed in earlier chapters on competition, movement, predation and parasitism. We must look at regulation, therefore, to understand how it is that a population tends to remain within defined upper and lower limits.

On the other hand, the precise abundance of individuals will be determined by the combined effects of all the processes that affect a population, whether they are dependent or independent of density. Figure 14.2 shows this diagrammatically and very simply.

Figure 14.2 (a) Population regulation with: (i) density-independent birth and density-dependent death; (ii) density-dependent birth and density-independent death; and (iii) density-dependent birth and death. Population size increases when the birth rate exceeds the death rate and decreases when the death rate exceeds the birth rate. N^* is therefore a stable equilibrium population size. The actual value of the equilibrium population size is seen to depend on both the magnitude of the density-independent rate and the magnitude and slope of any density-dependent process. (b) Population regulation with density-dependent birth, b , and density-independent death, d . Death rates are determined by physical conditions which differ in three sites (death rates d_1 , d_2 and d_3). Equilibrium population size varies as a result (N_1^* , N_2^* , N_3^*).



Here, the birth rate is density dependent, whilst the death rate is density independent but depends on physical conditions that differ in three locations. There are three equilibrium populations (N_1 , N_2 , N_3), which correspond to the three death rates, which in turn correspond to the physical conditions in the three environments. Variations in density-independent mortality like this were primarily responsible, for example, for differences in the abundance of the annual grass *Vulpia fasciculata* on different parts of a sand-dune environment in North Wales, UK. Reproduction was density dependent and regulatory, but varied little from site to site. However, physical conditions had strong density-independent effects on mortality (Watkinson & Harper, 1978). We must look at the determination of abundance, therefore, to understand how it is that a particular population exhibits a particular abundance at a particular time, and not some other abundance.

14.2.2 Theories of abundance

A. J. Nicholson The ‘stability’ viewpoint usually traces its roots back to A. J. Nicholson, a theoretical and laboratory animal ecologist working in Australia (e.g. Nicholson, 1954), believing that density-dependent, biotic interactions play the main role in determining population size, holding populations in a state of balance in their environments. Nicholson recognized, of course, that ‘factors which are uninfluenced by density may produce profound effects upon density’ (see Figure 14.2), but he considered that density dependence ‘is merely relaxed from time to time and subsequently resumed, and it remains the influence which adjusts population densities in relation to environmental favourability’.

The other point of view can be traced back to two other Australian ecologists, Andrewartha and Birch (1954), whose research was concerned mainly with the control of insect pests in the wild. It is likely, therefore, that their views were conditioned by the need to predict abundance and, especially, the timing and intensity of pest outbreaks. They believed that the most important factor limiting the numbers of organisms in natural populations was the shortage of time when the rate of increase in the population was positive. In other words, populations could be viewed as passing through a repeated sequence of setbacks and recovery – a view that can certainly be applied to many insect pests that are sensitive to unfavorable environmental conditions but are able to bounce back rapidly. They also rejected any subdivision of the environment into Nicholson’s density-dependent and density-independent ‘factors’, preferring instead to see populations as sitting at the center of an ecological web, where the essence was that various factors and processes interacted in their effects on the population.

With the benefit of hindsight, it seems clear that the first camp was preoccupied with what regulates population size and the second with what determines population size – and both are perfectly valid interests. Disagreement seems to have arisen because of some feeling within the first camp that whatever regulates *also* determines; and some feeling in the second camp that the determination of abundance is, for practical purposes, all that really matters. It is indisputable, however, that no population can be absolutely free of regulation – long-term unrestrained population growth is unknown, and

Andrewartha and Birch

no need for disagreement between the competing schools of thought

unrestrained declines to extinction are rare. Furthermore, any suggestion that density-dependent processes are rare or generally of only minor importance would be wrong. A very large number of studies have been made of various kinds of animals, especially of insects. Density dependence has by no means always been detected but is commonly seen when studies are continued for many generations. For instance, density dependence was detected in 80% or more of studies of insects that lasted more than 10 years (Hassell *et al.*, 1989; Woiwod & Hanski, 1992).

On the other hand, in the kind of study that Andrewartha and Birch focused on, weather was typically the major determinant of abundance and other factors were of relatively minor importance. For instance, in one famous, classic study of a pest, the apple thrips (*Thrips imaginis*), weather accounted for 78% of the variation in thrips numbers (Davidson & Andrewartha, 1948). To predict thrips abundance, information on the weather is of paramount importance. Hence, it is clearly not necessarily the case that whatever regulates the size of a population also determines its size for most of the time. And it would also be wrong to give regulation or density dependence some kind of preeminence. It may be occurring only infrequently or intermittently. And even when regulation is occurring, it may be drawing abundance toward a level that is itself changing in response to changing levels of resources. It is likely that no natural population is ever truly at equilibrium. Rather, it seems reasonable to expect to find some populations in nature that are almost always recovering from the last disaster (Figure 14.3a), others that are usually limited by an abundant resource (Figure 14.3b) or by a scarce resource (Figure 14.3c), and others that are usually in decline after sudden episodes of colonization (Figure 14.3d).

There is a very strong bias towards insects in the data sets available for the analysis of the regulation and determination of population size, and amongst these there is a preponderance of studies of pest species. The limited information from other groups suggests that terrestrial vertebrates may have significantly less variable populations than those of arthropods, and that populations of birds are more constant than those of mammals. Large terrestrial mammals seem to be regulated most often by their food supply, whereas in small mammals the single biggest cause of regulation seems to be the density-dependent exclusion of juveniles from breeding (Sinclair, 1989). For birds, food shortage and competition for territories and/or nest sites seem to be most important. Such generalizations, however, may be as much a reflection of biases in the species selected for study and of the neglect of their predators and parasites, as they are of any underlying pattern.

14.2.3 Approaches to the investigation of abundance

demographic,
mechanistic and
density approaches

Sibly and Hone (2002) distinguished three broad approaches that have been used to address questions about the determination and regulation of

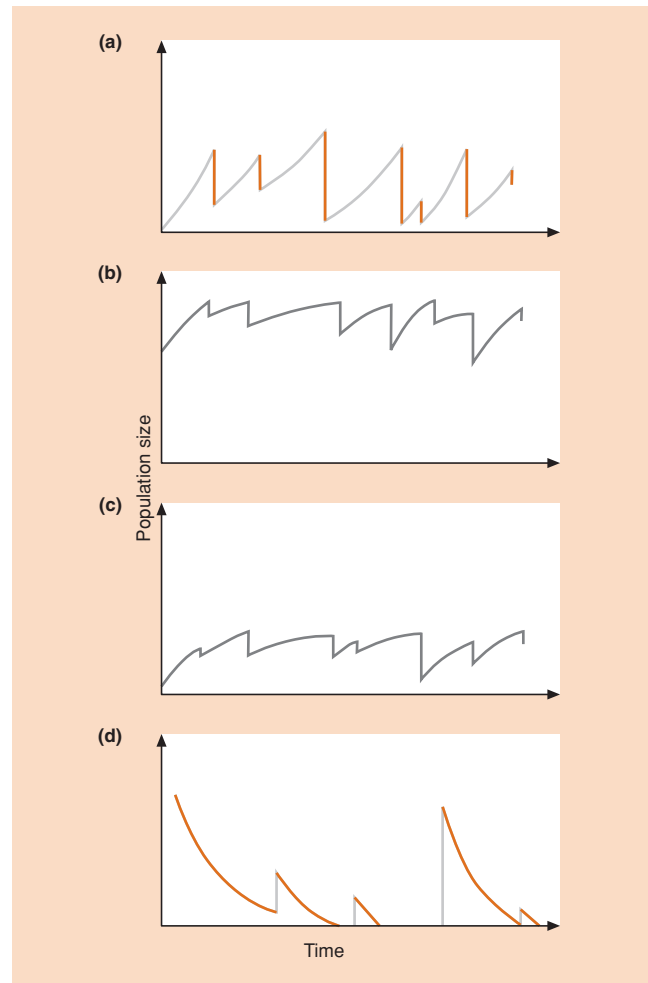


Figure 14.3 Idealized diagrams of population dynamics: (a) dynamics dominated by phases of population growth after disaster; (b) dynamics dominated by limitations on environmental carrying capacity – carrying capacity high; (c) same as (b) but carrying capacity low; and (d) dynamics within a habitable site dominated by population decay after more or less sudden episodes of colonization recruitment.

abundance. They did so having placed population growth rate at the center of the stage, since this summarizes the combined effects on abundance of birth, death and movement. The *demographic* approach (Section 14.3) seeks to partition variations in the overall population growth rate amongst the phases of survival, birth and movement occurring at different stages in the life cycle. The aim is to identify the most important phases. However, as we shall see, this begs the question ‘Most important for what?’ The *mechanistic* approach (Section 14.4) seeks to relate variations in growth rate directly to variations in specified factors – food, temperature, and so on – that might influence it. The approach itself can range from establishing correlations to carrying out field experiments. Finally, the *density* approach (Section 14.5) seeks to

relate variations in growth rate to variations in density. This is a convenient framework for us to use in examining some of the wide variety of studies that have been carried out. However, as Sibly and Hone’s (2002) survey makes clear, many studies are hybrids of two, or even all three, of the approaches. Lack of space will prevent us from looking at all of the different variants.

14.3 The demographic approach

14.3.1 Key factor analysis

key factors? or key phases?

For many years, the demographic approach was represented by a technique called *key factor analysis*. As we shall see, there are shortcomings in the technique and useful modifications have been proposed, but as a means of explaining important general principles, and for historical completeness, we start with key factor analysis. In fact, the technique is poorly named, since it begins, at least, by identifying key *phases* (rather than factors) in the life of the organism concerned.

the Colorado potato beetle

For a key factor analysis, data are required in the form of a series of life tables (see Section 4.5) from a number of different cohorts of the population concerned. Thus, since its initial development (Morris, 1959; Varley & Gradwell, 1968) it has been most commonly used for species with discrete generations, or where cohorts can otherwise be readily distinguished. In particular, it is an approach based on the use of *k* values (see Sections 4.5.1 and 5.6). An example, for a Canadian population of the Colorado potato beetle (*Leptinotarsa decemlineata*), is shown in Table 14.1 (Harcourt, 1971). In this species, ‘spring adults’ emerge from hibernation around the middle of June, when potato plants are breaking through the ground. Within 3

or 4 days oviposition (egg laying) begins, continuing for about 1 month and reaching its peak in early July. The eggs are laid in clusters on the lower leaf surface, and the larvae crawl to the top of the plant where they feed throughout their development, passing through four instars. When mature, they drop to the ground and pupate in the soil. The ‘summer adults’ emerge in early August, feed, and then re-enter the soil at the beginning of September to hibernate and become the next season’s ‘spring adults’.

The sampling program provided estimates of the population at seven stages: eggs, early larvae, late larvae, pupae, summer adults, hibernating adults and spring adults. One further category was included, ‘females × 2’, to take account of any unequal sex ratios amongst the summer adults. Table 14.1 lists these estimates for a single season. It also gives what were believed to be the main causes of death in each stage of the life cycle. In so doing, what is essentially a demographic technique (dealing with phases) takes on the mantle of a mechanistic approach (by associating each phase with a proposed ‘factor’).

mean *k* values: typical strengths of factors

The mean *k* values, determined for a single population over 10 seasons, are presented in the third column of Table 14.2. These indicate the relative strengths of the various factors that contribute to the total rate of mortality within a generation. Thus, the emigration of summer adults has by far the greatest proportional effect ($k_6 = 1.543$), whilst the starvation of older larvae, the frost-induced mortality of hibernating adults, the ‘nondeposition’ of eggs, the effects of rainfall on young larvae and the cannibalization of eggs all play substantial roles as well.

What this column of Table 14.2 does not tell us, however, is the relative importance of these factors as determinants of the year-to-year *fluctuations* in mortality. For instance, we can easily imagine a factor that repeatedly takes a significant toll from a population, but which, by remaining constant in its effects, plays

Table 14.1 Typical set of life table data collected by Harcourt (1971) for the Colorado potato beetle (in this case for Merivale, Canada, 1961–62).

Age interval	Numbers per 96		'Mortality factor'	Log ₁₀ N	k value
	potato hills	Numbers 'dying'			
Eggs	11,799	2,531	Not deposited	4.072	0.105 (k_{1a})
	9,268	445	Infertile	3.967	0.021 (k_{1b})
	8,823	408	Rainfall	3.946	0.021 (k_{1c})
	8,415	1,147	Cannibalism	3.925	0.064 (k_{1d})
	7,268	376	Predators	3.861	0.024 (k_{1e})
Early larvae	6,892	0	Rainfall	3.838	0 (k_2)
Late larvae	6,892	3,722	Starvation	3.838	0.337 (k_3)
Pupal cells	3,170	16	<i>D. doryphorae</i>	3.501	0.002 (k_4)
Summer adults	3,154	126	Sex (52% ♀)	3.499	-0.017 (k_5)
♀ × 2	3,280	3,264	Emigration	3.516	2.312 (k_6)
Hibernating adults	16	2	Frost	1.204	0.058 (k_7)
Spring adults	14			1.146	
					2.926 (k_{total})

Mortality factor	k	Mean k value	Regression coefficient on k_{total}	b	r^2
Eggs not deposited	k_{1a}	0.095	-0.020	-0.05	0.27
Eggs infertile	k_{1b}	0.026	-0.005	-0.01	0.86
Rainfall on eggs	k_{1c}	0.006	0.000	0.00	0.00
Eggs cannibalized	k_{1d}	0.090	-0.002	-0.01	0.02
Eggs predation	k_{1c}	0.036	-0.011	-0.03	0.41
Larvae 1 (rainfall)	k_2	0.091	0.010	0.03	0.05
Larvae 2 (starvation)	k_3	0.185	0.136	0.37	0.66
Pupae (<i>D. doryphorae</i>)	k_4	0.033	-0.029	-0.11	0.83
Unequal sex ratio	k_5	-0.012	0.004	0.01	0.04
Emigration	k_6	1.543	0.906	2.65	0.89
Frost	k_7	0.170	0.010	0.002	0.02
	k_{total}	2.263			

Table 14.2 Summary of the life table analysis for Canadian Colorado beetle populations. b is the slope of the regression of each k factor on the logarithm of the numbers preceding its action; r^2 is the coefficient of determination. See text for further explanation. (After Harcourt, 1971.)

little part in determining the particular rate of mortality (and thus, the particular population size) in any 1 year. This can be assessed, however, from the next column of Table 14.2, which gives the regression coefficient of each individual k value on the total generation value, k_{total} .

regressions of k on k_{total} : key factors

A mortality factor that is important in determining population changes will have a regression coefficient close to unity, because its k value will tend to fluctuate in line with k_{total} in terms of both size and direction (Podoler & Rogers, 1975). A mortality factor with a k value that varies quite randomly with respect to k_{total} , however, will have a regression coefficient close to zero. Moreover, the sum of all the regression coefficients within a generation will always be unity. The values of the regression coefficients will, therefore, indicate the relative strength of the association between different factors and the fluctuations in mortality. The largest regression coefficient will be associated with the *key factor* causing population change.

In the present example, it is clear that the emigration of summer adults, with a regression coefficient of 0.906, is the key factor. Other factors (with the possible exception of larval starvation) have a negligible effect on the changes in generation mortality, even though some have reasonably high mean k values. A similar conclusion can be drawn by simply examining graphs of the fluctuations in k values with time (Figure 14.4a).

Thus, whilst mean k values indicate the average strengths of various factors as causes of mortality in each generation, key factor analysis indicates their relative contribution to the yearly changes in generation mortality, and thus measures their importance as determinants of population size.

a role for factors in regulation?

What, though, of population regulation? To address this, we examine the density dependence of each factor by plotting k values against \log_{10} of the

numbers present before the factor acted (see Section 5.6). Thus, the last two columns in Table 14.2 contain the slopes (b) and coefficients of determination (r^2) of the various regressions of k values on their appropriate 'log₁₀ initial densities'. Three factors seem worthy of close examination. The emigration of summer adults (the key factor) appears to act in an overcompensating density-dependent fashion, since the slope of the regression (2.65) is considerably in excess of unity (see also Figure 14.4b). Thus, the key factor, although density dependent, does not so much regulate the population as lead to violent fluctuations in abundance (because of overcompensation). Indeed, the Colorado potato beetle-potato system would go extinct if potatoes were not continually replanted (Harcourt, 1971).

Also, the rate of larval starvation appears to exhibit undercompensating density dependence (although statistically this is not significant). An examination of Figure 14.4b, however, shows that the relationship would be far better represented not by a linear regression but by a curve. If such a curve is fitted to the data, then the coefficient of determination rises from 0.66 to 0.97, and the slope (b value) achieved at high densities would be 30.95 (although it is, of course, much less than this in the range of densities observed). Hence, it is quite possible that larval starvation plays an important part in regulating the population, prior to the destabilizing effects of pupal parasitism and adult emigration.

Key factor analysis has been applied to a great many insect populations, but to far fewer vertebrate or plant populations. Examples of these, though, are shown in Table 14.3 and Figure 14.5. In populations of the wood frog (*Rana sylvatica*) in three regions of the United States (Table 14.3), the larval period was the key phase determining abundance in each region (second data column), largely as a result of year-to-year variations in rainfall during the larval period. In low rainfall years, the ponds could dry out, reducing larval survival to catastrophic levels, sometimes as a result of a bacterial infection. Such

wood frogs and an annual plant

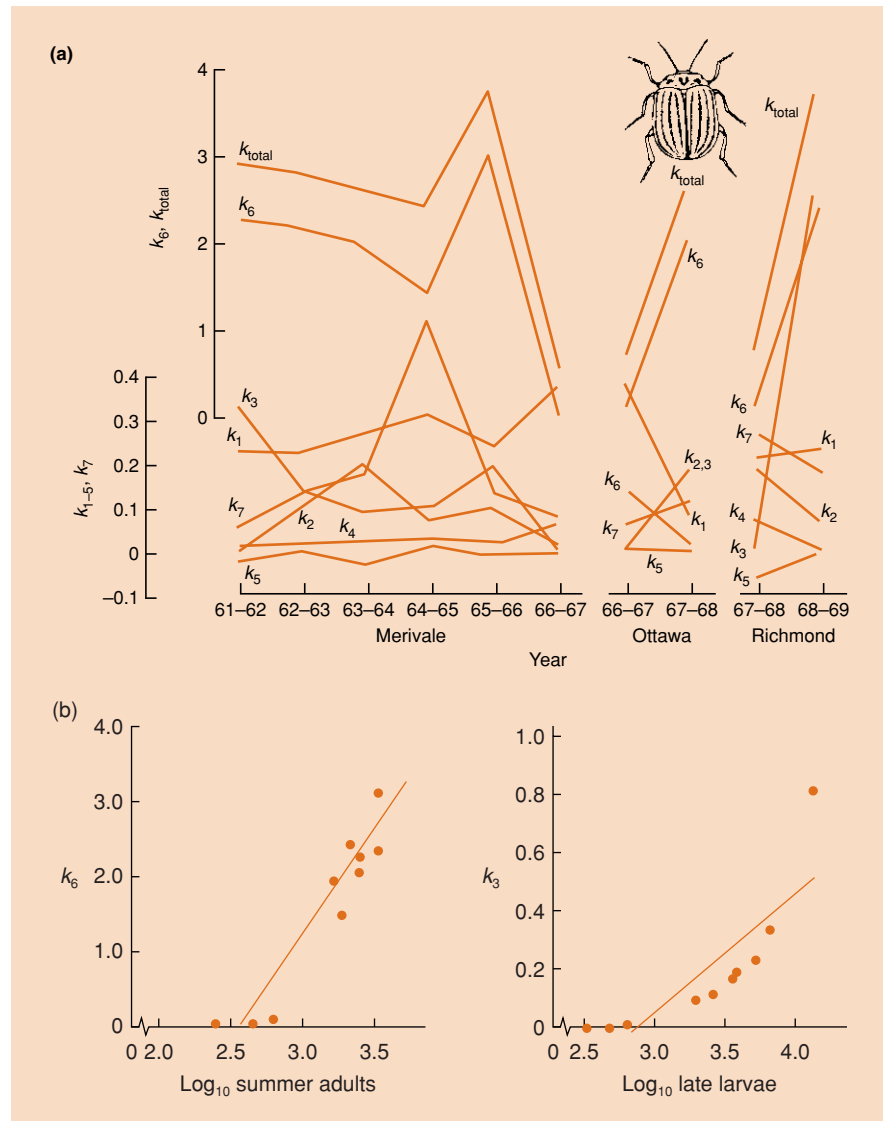


Figure 14.4 (a) The changes with time of the various k values of Colorado beetle populations at three sites in Canada. (After Harcourt, 1971.) (b) Density-dependent emigration by Colorado beetle ‘summer’ adults (slope = 2.65) (left) and density-dependent starvation of larvae (slope = 0.37) (right). (After Harcourt, 1971.)

mortality, however, was inconsistently related to the size of the larval population (one pond in Maryland, and only approaching significance in Virginia – third data column) and hence played an inconsistent part in regulating the sizes of the populations. Rather, in two of the regions it was during the adult phase that mortality was clearly density dependent and hence regulatory (apparently as a result of competition for food). Indeed, in two of the regions mortality was also most intense in the adult phase (first data column).

The key phase determining abundance in a Polish population of the sand-dune annual plant *Androsace septentrionalis* (Figure 14.5; see also Figure 14.1a) was found to be the seeds in the soil. Once

again, however, mortality did not operate in a density-dependent manner, whereas mortality of seedlings, which were not the key phase, was found to be density dependent. Seedlings that emerge first in the season stand a much greater chance of surviving.

Overall, therefore, key factor analysis (its rather misleading name apart) is useful in identifying important phases in the life cycles of study organisms. It is useful too in distinguishing the variety of ways in which phases may be important: in contributing significantly to the overall sum of mortality; in contributing significantly to variations in mortality, and hence in determining abundance; and in contributing significantly to the regulation of abundance by virtue of the density dependence of the mortality.

Age interval	Mean k value	Coefficient of regression on k_{total}	Coefficient of regression on $\log(\text{population size})$
Maryland			
Larval period	1.94	0.85	Pond 1: 1.03 ($P = 0.04$) Pond 2: 0.39 ($P = 0.50$)
Juvenile: up to 1 year	0.49	0.05	0.12 ($P = 0.50$)
Adult: 1–3 years	2.35	0.10	0.11 ($P = 0.46$)
Total	4.78		
Virginia			
Larval period	2.35	0.73	0.58 ($P = 0.09$)
Juvenile: up to 1 year	1.10	0.05	-0.20 ($P = 0.46$)
Adult: 1–3 years	1.14	0.22	0.26 ($P = 0.05$)
Total	4.59		
Michigan			
Larval period	1.12	1.40	1.18 ($P = 0.33$)
Juvenile: up to 1 year	0.64	1.02	0.01 ($P = 0.96$)
Adult: 1–3 years	3.45	-1.42	0.18 ($P = 0.005$)
Total	5.21		

Table 14.3 Key factor (or key phase) analysis for wood frog populations from three areas in the United States: Maryland (two ponds, 1977–82), Virginia (seven ponds, 1976–82) and Michigan (one pond, 1980–93). In each area, the phase with the highest mean k value, the key phase and any phase showing density dependence are highlighted in bold. (After Berven, 1995.)

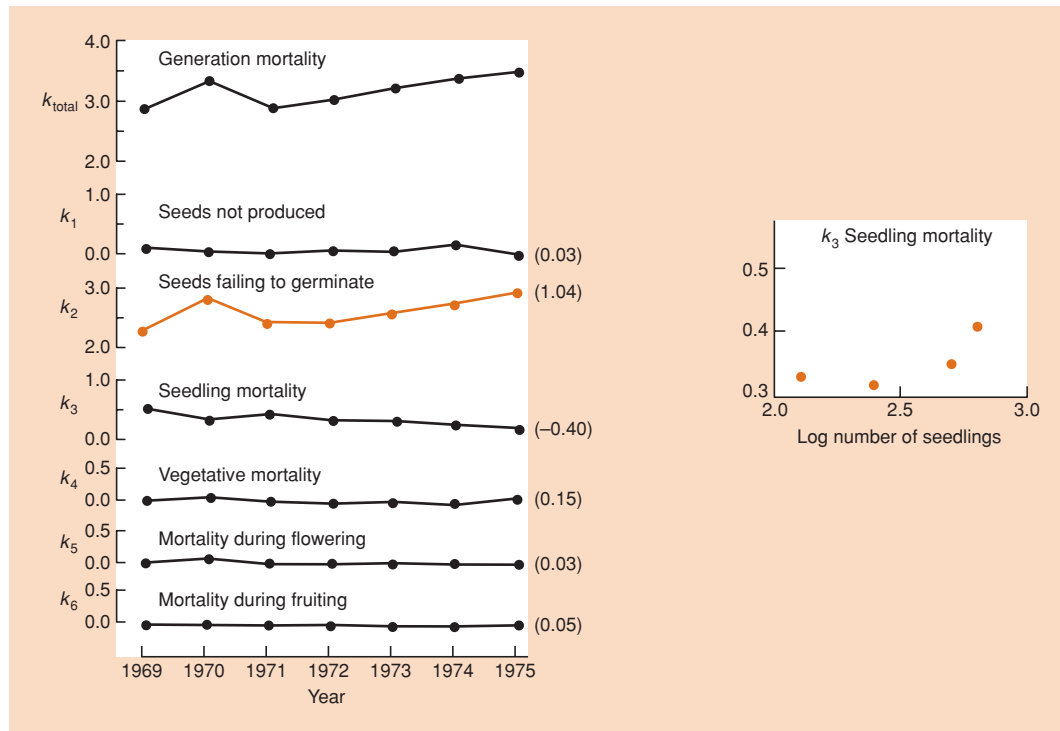


Figure 14.5 Key factor analysis of the sand-dune annual plant *Androsace septentrionalis*. A graph of total generation mortality (k_{total}) and of various k factors is presented. The values of the regression coefficients of each individual k value on k_{total} are given in brackets. The largest regression coefficient signifies the key phase and is shown as a colored line. Alongside is shown the one k value that varies in a density-dependent manner. (After Symonides, 1979; analysis in Silvertown, 1982.)

14.3.2 Sensitivities, elasticities and λ -contribution analysis

overcoming problems in key factor analysis

Although key factor analysis has been useful and widely used, it has been subject to persistent and valid criticisms, some technical (i.e. statistical) and some conceptual (Sibly & Smith, 1998). Important among these criticisms are: (i) the rather awkward way in which k values deal with fecundity: a value is calculated for ‘missing’ births, relative to the maximum possible number of births; and (ii) ‘importance’ may be inappropriately ascribed to different phases, because equal weight is given to all phases of the life history, even though they may differ in their power to influence abundance. This is a particular problem for populations in which the generations overlap, since mortalities (and fecundities) later in the life cycle are bound to have less effect on the overall rate of population growth than those occurring in earlier phases. In fact, key factor analysis was designed for species with discrete generations, but it has been applied to species with overlapping generations, and in any case, restricting it to the former is a limitation on its utility.

Sibly and Smith’s (1998) alternative to key factor analysis, λ -contribution analysis, overcomes these problems. λ is the population growth rate (e^r) that we referred to as R , for example, in Chapter 4, but here we retain Sibly and Smith’s notation. Their method, in turn, makes use of a weighting of life cycle phases taken from sensitivity and elasticity analysis (De Kroon *et al.*, 1986; Benton & Grant, 1999; Caswell, 2001; see also ‘integral projection models’, for example Childs *et al.* (2003)), which is itself an important aspect of the demographic approach to the study of abundance. Hence, we deal first, briefly, with sensitivity and elasticity analysis before examining λ -contribution analysis.

the population projection matrix revisited

The details of calculating sensitivities and elasticities are beyond our scope, but the principles can best be understood by returning to the population projection matrix, introduced in Section 4.7.3.

Remember that the birth and survival processes in a population can be summarized in matrix form as follows:

$$\begin{bmatrix} p_0 & m_1 & m_2 & m_3 \\ g_0 & p_1 & 0 & 0 \\ 0 & g_1 & p_2 & 0 \\ 0 & 0 & g_2 & p_3 \end{bmatrix}$$

where, for each time step, m_x is the fecundity of stage x (into the first stage), g_x is the rate of survival and growth from stage x into the next stage, and p_x is the rate of persisting within stage x . Remember, too, that λ can be computed directly from this matrix. Clearly, the overall value of λ reflects the values of the various elements in the matrix, but their contribution to λ is not equal. The *sensitivity*, then, of each element (i.e. each biological process)

is the amount by which λ would change for a given absolute change in the value of the matrix element, with the value of all the other elements held constant. Thus, sensitivities are highest for those processes that have the greatest power to influence λ .

However, whereas survival elements (g s and p s here) are constrained to lie between 0 and 1, fecundities are not, and λ therefore tends to be more sensitive to absolute changes in survival than to absolute changes of the same magnitude in fecundity. Moreover, λ can be sensitive to an element in the matrix even if that element takes the value 0 (because sensitivities measure what *would* happen if there *was* an absolute change in its value). These shortcomings are overcome, though, by using the *elasticity* of each element to determine its contribution to λ , since this measures the proportional change in λ resulting from a proportional change in that element. Conveniently, too, with this matrix formulation the elasticities sum to 1.

sensitivity and elasticity

Elasticity analysis therefore offers an especially direct route to plans for the management of abundance. If we wish to increase the abundance of a threatened species (ensure λ is as high as possible) or decrease the abundance of a pest (ensure λ is as low as possible), which phases in the life cycle should be the focus of our efforts? Answer: those with the highest elasticities. For example, an elasticity analysis of the threatened Kemp’s ridley sea turtle (*Lepidochelys kempi*) off the southern United States showed that the survival of older, especially subadult individuals was more critical to the maintenance of abundance than either fecundity or hatchling survival (Figure 14.6a). Therefore, ‘headstarting’ programs, in which eggs were reared elsewhere (Mexico) and imported, and which had dominated conservation practice through the 1980s, seem doomed to be a low-payback management option (Heppell *et al.*, 1996). Worryingly, headstarting programs have been widespread, and yet this conclusion seems likely to apply to turtles generally.

elasticity analysis and the management of abundance

Elasticity analysis was applied, too, to populations of the nodding thistle (*Carduus nutans*), a noxious weed in New Zealand. The survival and reproduction of young plants were far more important to the overall population growth rate than those of older individuals (Figure 14.6b), but, discouragingly, although the bio-control program in New Zealand had correctly targeted these phases through the introduction of the seed-eating weevil, *Rhinocyllus conicus*, the maximum observed levels of seed predation (c. 49%) were lower than those projected to be necessary to bring λ below 1 (69%) (Shea & Kelly, 1998). As predicted, the control program has had only limited success.

Thus, elasticity analyses are valuable in identifying phases and processes that are important in determining abundance, but they do so by focusing on typical or average values, and in that

elasticity may say little about variations in abundance ...

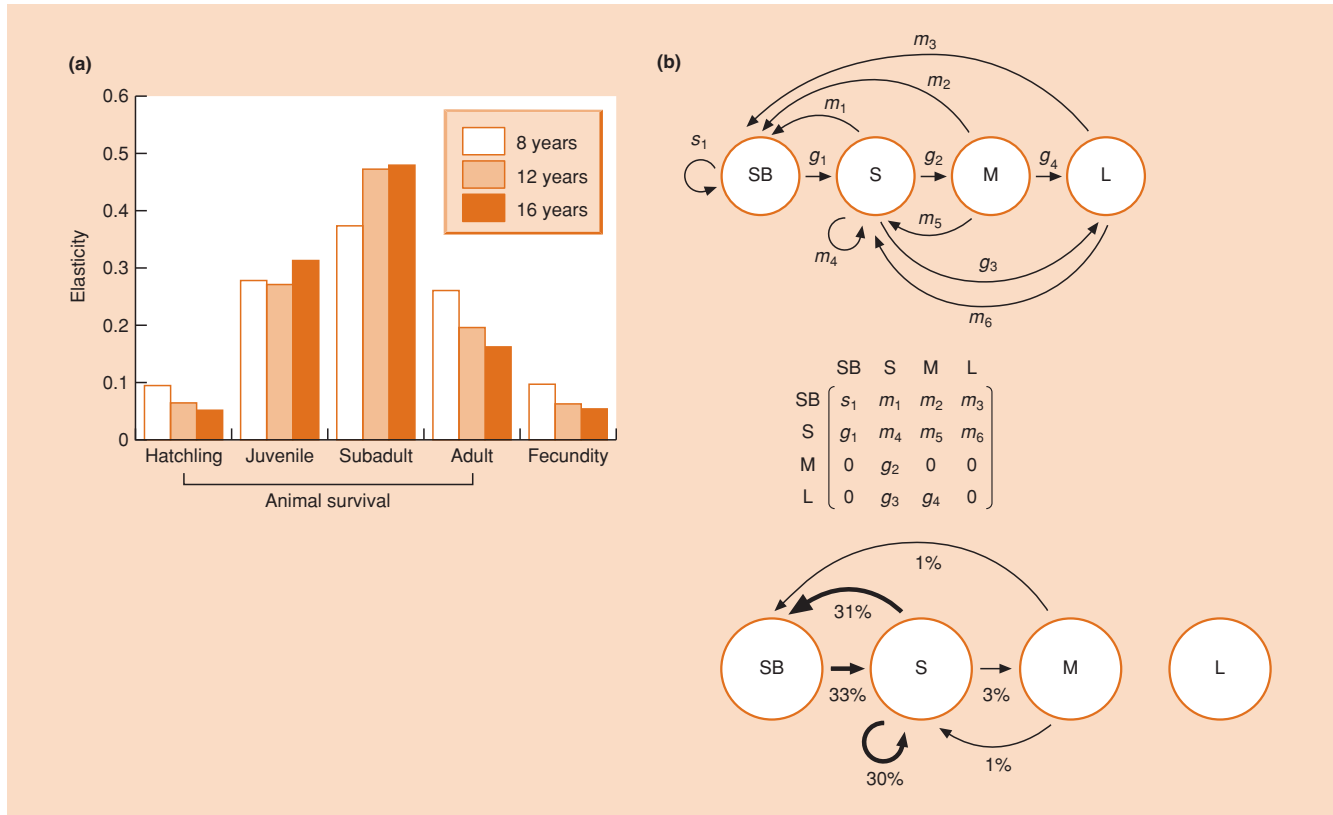


Figure 14.6 (a) Results of elasticity analyses for Kemp’s ridley turtles (*Lepidochelys kempi*), showing the proportional changes in λ resulting from proportional changes in stage-specific annual survival and fecundity, on the assumption of three different ages of maturity. (After Heppell *et al.*, 1996.) (b) Top: diagrammatic representation of the life cycle structure of *Carduus nutans* in New Zealand, where SB is the seed bank and S, M and L are small, medium and large plants, and s is seed dormancy, g is growth and survival to subsequent stages, and m is the reproductive contribution either to the seed bank or to immediately germinating small plants. Middle: the population projection matrix summarizing this structure. Bottom: the results of an elasticity analysis for one population, in which the percentage changes in λ resulting from percentage changes in s , g and r are shown on the life cycle diagram. The most important transitions are shown in bold, and elasticities less than 1% are omitted altogether. (After Shea & Kelly, 1998.)

sense they seek to account for the typical size of a population. However, a process with a high elasticity may still play little part, in practice, in accounting for variations in abundance from year to year or site to site, if that process (mortality or fecundity) shows little temporal or spatial variation. There is even evidence from large herbivorous mammals that processes with high elasticity tend to vary little over time (e.g. adult female fecundity), whereas those with low elasticity (e.g. juvenile survival) vary far more (Gaillard *et al.*, 2000). The *actual* influence of a process on variations in abundance will depend on both elasticity *and* variation in the process. Gaillard *et al.* further suggest that the relative absence of variation in the ‘important’ processes may be a case of ‘environmental canalisation’: evolution, in the phases most important to fitness, of an ability to maintain relative constancy in the face of environmental perturbations.

In contrast to elasticity analyses, key factor analysis seeks specifically to understand temporal and spatial variations in abundance. The same is true of Sibly and Smith’s λ -contribution analysis, to which we now return. We can note first that it deals with the contributions of the different phases not to an overall k value (as in key factor analysis) but to λ , a much more obvious determinant of abundance. It makes use of k values to quantify mortality, but can use fecundities directly rather than converting them into ‘deaths of unborn offspring’. And crucially, the contributions of all mortalities and fecundities are weighted by their sensitivities. Hence, quite properly, where generations overlap, the chances of later phases being identified with a key factor are correspondingly lower in λ -contribution than in key factor

... but λ -contribution analysis does

analysis. As a result, λ -contribution analysis can be used with far more confidence when generations overlap. Subsequent investigation of density dependences proceeds in exactly the same way in λ -contribution analysis as in key factor analysis.

Table 14.4 contrasts the results of the two analyses applied to life table data collected on the Scottish island of Rhum between 1971 and 1983 for the red deer, *Cervus elaphus* (Clutton-Brock *et al.*, 1985). Over the 19-year lifespan of the deer, survival and birth rates were estimated in the following 'blocks': year 0, years 1 and 2, years 3 and 4, years 5–7, years 8–12 and years 13–19. This accounts for the limited number of different values in the k_x and m_x columns of the table, but the sensitivities of λ to these values are of course different for different ages (early influences on λ are more powerful), with the exception that λ is equally sensitive to mortality in each phase prior to first reproduction (since it is all 'death before reproduction'). The consequences of these differential sensitivities are apparent in the final two columns

of the table, which summarize the results of the two analyses by presenting the regression coefficients of each of the phases against k_{total} and λ_{total} , respectively. Key factor analysis identifies reproduction in the final years of life as the key factor and even identifies reproduction in the preceding years as the next most important phase. In stark contrast, in λ -contribution analysis, the low sensitivities of λ to birth in these late phases relegate them to relative insignificance – especially the last phase. Instead, survival in the earliest phase of life, where sensitivity is greatest, becomes the key factor, followed by fecundity in the 'middle years' where fecundity itself is highest. Thus, λ -contribution analysis combines the virtues of key factor and elasticity analyses: distinguishing the regulation and determination of abundance, identifying key phases or factors, while taking account of the differential sensitivities of growth rate (and hence abundance) to the different phases.

Table 14.4 Columns 1–4 contain life table data for the females of a population of red deer, *Cervus elaphus*, on the island of Rhum, Scotland, using data collected between 1971 and 1983 (Clutton-Brock *et al.*, 1985): x is age, l_x is the proportion surviving at the start of an age class, k_x , killing power, has been calculated using natural logarithms, and m_x , fecundity, refers to the birth of female calves. These data represent averages calculated over the period, the raw data having been collected both by following individually recognizable animals from birth and aging animals at death. The next two columns contain the sensitivities of λ , the population growth rate, to k_x and m_x in each age class. In the final two columns, the contributions of the various age classes have been grouped as shown. These columns show the contrasting results of a key factor analysis and a λ contribution analysis as the regression coefficients of k_x and m_x on k_{total} and λ_{total} , respectively, where λ_{total} is the deviation each year from the long-term average value of λ . (After Sibly & Smith, 1998, where details of the calculations may also be found.)

Age (years) at start of class, x	l_x	k_x	m_x	Sensitivity of λ to k_x	Sensitivity of λ to m_x	Regression coefficients of k_x , left, and m_x , right, on k_{total}	Regression coefficients of k_x , left, and m_x , right, on λ_{total}
0	1.00	0.45	0.00	-0.14	0.16	0.01, -	0.32, -
1	0.64	0.08	0.00	-0.14	0.09	0.01, -	0.14, -
2	0.59	0.08	0.00	-0.14	0.08		
3	0.54	0.03	0.22	-0.13	0.07	0.00, 0.05	0.03, 0.04
4	0.53	0.03	0.22	-0.11	0.06		
5	0.51	0.04	0.35	-0.10	0.05	-0.00, 0.03	0.08, 0.16
6	0.49	0.04	0.35	-0.08	0.05		
7	0.47	0.04	0.35	-0.07	0.04		
8	0.45	0.06	0.37	-0.05	0.04	0.01, 0.15	0.09, 0.12
9	0.42	0.06	0.37	-0.04	0.03		
10	0.40	0.06	0.37	-0.03	0.03		
11	0.38	0.06	0.37	-0.02	0.02		
12	0.35	0.06	0.37	-0.02	0.02	-0.05, 0.80	0.01, -0.00
13	0.33	0.30	0.30	-0.01	0.02		
14	0.25	0.30	0.30	-0.006	0.01		
15	0.18	0.30	0.30	-0.004	0.008		
16	0.14	0.30	0.30	-0.002	0.005		
17	0.10	0.30	0.30	-0.001	0.004		
18	0.07	0.30	0.30	-0.001	0.002		
19	0.06	0.30	0.30	-0.000	0.002		

14.4 The mechanistic approach

The previous section dealt with analyses directed at phases in the life cycle, but these often ascribe the effects occurring in particular phases to factors or processes – food, predation, etc. – known to operate during those phases. An alternative has been to study the role of particular factors in the determination of abundance directly, by relating the level or presence of the factor (the amount of food, the presence of predators) either to abundance itself or to population growth rate, which is obviously the proximate determinant of abundance. This mechanistic approach has the advantage of focusing clearly on the particular factor, but in so doing it is easy to lose sight of the relative importance of that factor compared to others.

14.4.1 Correlating abundance with its determinants

Figure 14.7, for example, shows four examples in which population growth rate increases with the availability of food. It also suggests that in general, such relationships are likely to level off at the highest food levels where some other factor or factors place an upper limit on abundance.

14.4.2 Experimental perturbation of populations

As we noted in the introduction to this chapter, correlations can be suggestive, but a much more powerful test of the importance

of a particular factor is to manipulate that factor and monitor the population's response. Predators, competitors or food can be added or removed, and if they are important in determining abundance, this should be apparent in subsequent comparisons of control and manipulated populations. Examples are discussed below, when we examine what may drive the regular cycles of abundance exhibited by some species, but we should note straight away that field-scale experiments require major investments in time and effort (and money), and a clear distinction between controls and experimental treatments is inevitably much more difficult to achieve than in the laboratory or greenhouse.

One context in which predators have been added to a population is when biological control agents (natural enemies of a pest – see Section 15.2.5) have been released in attempts to control pests. However, because the motivation has been practical rather than intellectual, perfect experimental design has not usually been a priority. There have, for example, been many occasions when aquatic plants have undergone massive population explosions after their introduction to new habitats, creating significant economic problems by blocking navigation channels and irrigation pumps and upsetting local fisheries. The population explosions occur as the plants grow clonally, break up into fragments and become dispersed. The aquatic fern, *Salvinia molesta*, for instance, which originated in southeastern Brazil, has appeared since 1930 in various tropical and subtropical regions. It was first recorded in Australia in 1952 and spread very rapidly – under optimal conditions *Salvinia* has a doubling time of 2.5 days.

biological control:
an experimental
perturbation

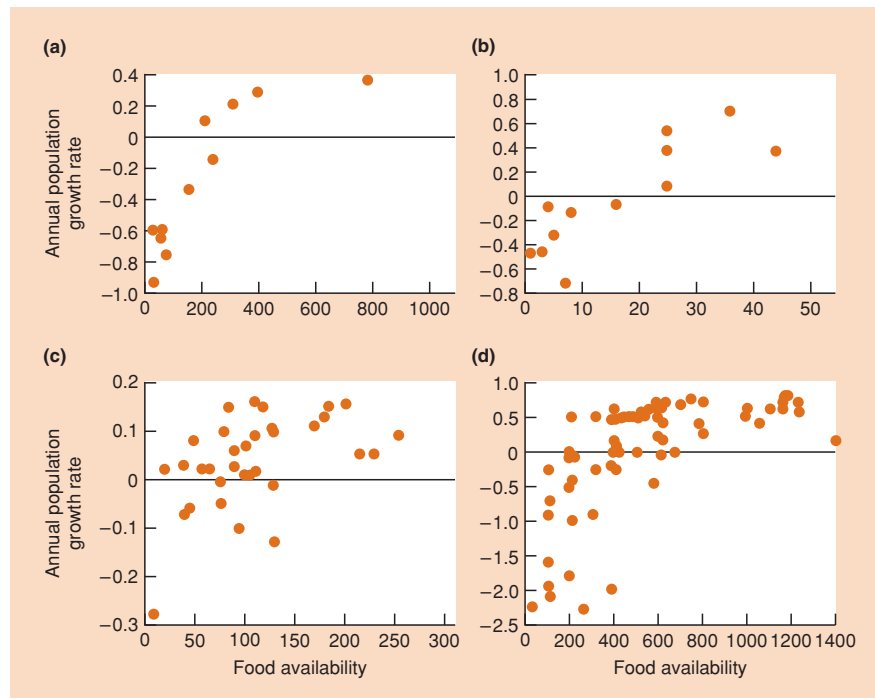


Figure 14.7 Increases in annual population growth rate ($r = \ln \lambda$) with the availability of food (pasture biomass (in kg ha^{-1}), except in (b) where it is vole abundance and in (c) where it is availability per capita). (a) Red kangaroo (from Bayliss, 1987). (b) Barn owl (after Taylor, 1994). (c) Wildebeest (from Krebs *et al.*, 1999). (d) Feral pig (from Choquenot, 1998). (After Sibly & Hone, 2002.)

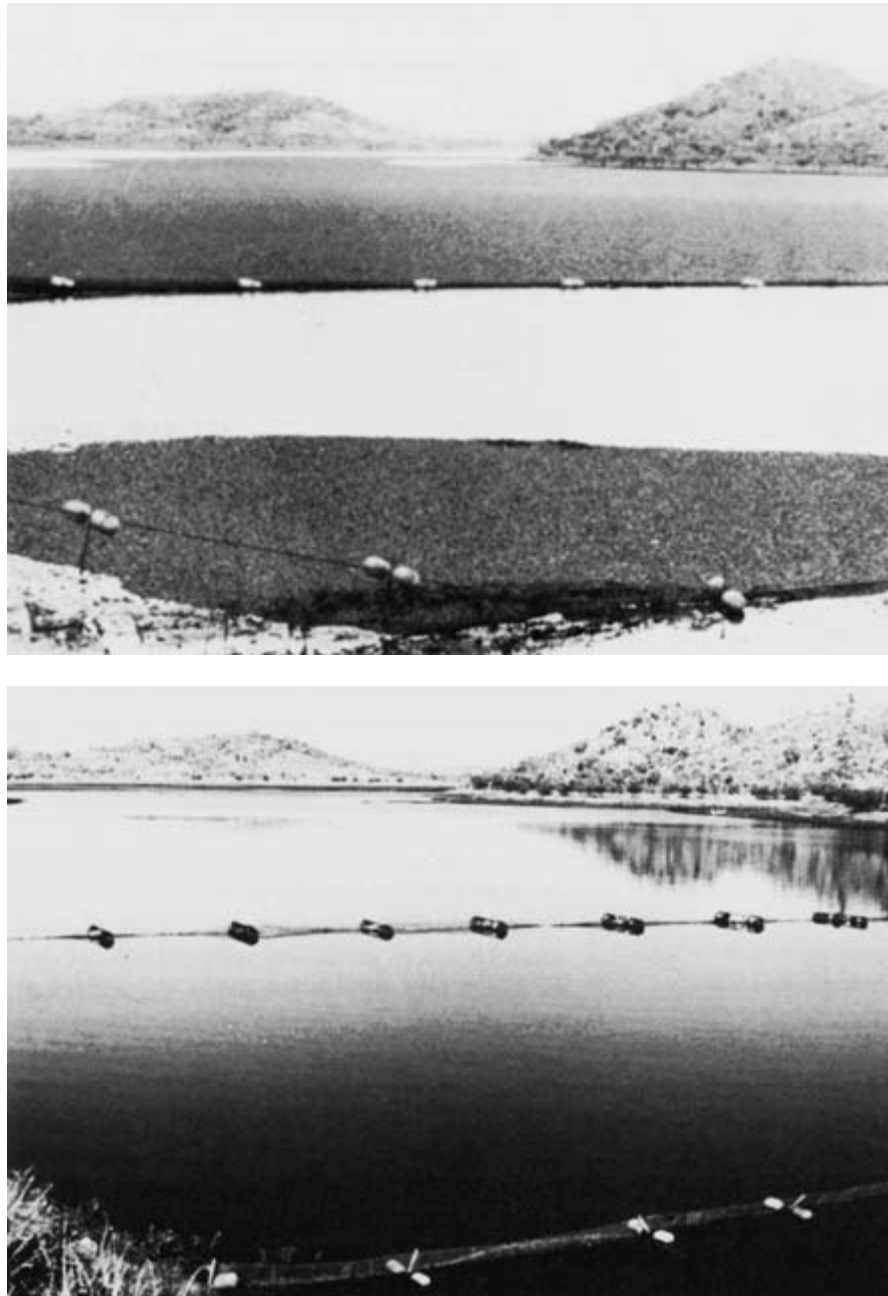


Figure 14.8 Lake Moon Darra (North Queensland, Australia). (a) Covered by dense populations of the water fern (*Salvinia molesta*). (b) After the introduction of weevil (*Cyrtobagous* spp.). (Courtesy of P.M. Room.)

Significant pests and parasites appear to have been absent. In 1978, Lake Moon Darra (northern Queensland) carried an infestation of 50,000 tonnes fresh weight of *Salvinia* covering an area of 400 ha (Figure 14.8).

Amongst the possible control agents collected from *Salvinia*'s native range in Brazil, the black long-snouted weevil (*Cyrtobagous* sp.) was known to feed only on *Salvinia*. On June 3, 1980, 1500 adults were released in cages at an inlet to the lake and a further release was made on January 20, 1981. The weevil was free of any parasites or predators that might reduce its density and,

by April 1981, *Salvinia* throughout the lake had become dark brown. Samples of the dying weed contained around 70 adult weevils per square meter, suggesting a total population of 1000 million beetles on the lake. By August 1981, there was estimated to be less than 1 tonne of *Salvinia* left on the lake (Room *et al.*, 1981). This has been the most rapid success of any attempted biological control of one organism by the introduction of another, and it establishes the importance of the weevil in the persistently low abundance of *Salvinia* both after the weevil's introduction to Australia and in its native environment. It was a controlled

experiment to the extent that other lakes continued to bear large populations of *Salvinia*.

treating red grouse
for nematodes

Both the power and the problems of field-scale experiments are further illustrated by an example we have already discussed in Section 12.7.2, in which a ‘predator’ (in this case a parasite) was not added but removed. When Hudson *et al.* (1998) treated cyclic populations of the red grouse *Lagopus lagopus* against the nematode *Trichostrongylus tenuis*, the extent of the grouse ‘crash’ was very substantially reduced, proving the importance of the nematodes, normally, in reducing grouse abundance, and justifying the effort that had gone into the manipulation. But as we have seen, in spite of this effort, controversy remained about whether the nematodes had been proved to be the cause of the cycles (in which case, the residual smaller crashes were dying echoes) or whether, instead, the experiment had only proved a role for the nematodes in determining a cycle’s amplitude, leaving their role in cyclicity itself uncertain. Experiments are better than correlations, but when they involve ecological systems in the field, eliminating ambiguity can never be guaranteed.

14.5 The density approach

Correlations with density have not been altogether absent from the approaches we have considered so far, and indeed, density dependence played a central role in our discussions of the determinants of abundance (birth, death and movement) in earlier chapters. Some studies, however, have focused much more fixedly on density dependences in their own right. In particular, many such studies have been designed to seek evidence for both direct and *delayed* density dependence (see Section 10.2.2). It is a problem, for example, that conventional life table analyses may fail to detect delayed density dependence simply because they are not designed to do so (Turchin, 1990). An analysis of population time series for 14 species of forest insects detected direct density dependence clearly in only five, but it revealed delayed density dependence in seven of the remaining nine (Turchin, 1990). It may be that a similar proportion of populations classified, from their life tables, as lacking density dependence are actually subject to the delayed density dependence of a natural enemy.

14.5.1 Time series analysis: dissecting density dependence

abundance
determination
expressed as a
time-lag equation

A number of related approaches have sought to dissect the density-dependent ‘structure’ of species’ population dynamics by a statistical analysis of time series of abundance. Abundance at

a given point in time may be seen as reflecting abundances at various times in the past. It reflects abundance in the immediate past in the obvious sense that the past abundance gave rise directly to the present abundance. It may also reflect abundance in the more distant past if, for example, that past abundance gave rise to an increased abundance of a predator, which in due course affected the present abundance (i.e. a delayed density dependence). In particular, and without going into technical details, the log of the abundance of a population at time t , X_t , can be expressed, at least approximately, as:

$$X_t = m + (1 + \beta_1)X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_d X_{t-d} + u_t, \quad (14.1)$$

an equation that captures, in a particular functional form, the idea of present abundance being determined by past abundances (Royama, 1992; Bjørnstad *et al.*, 1995; see also Turchin & Berryman, 2000). Thus, m reflects the mean abundance around which there are fluctuations over time, β_1 reflects the strength of direct density dependence, and other β s reflect the strengths of delayed density dependences with various time lags up to a maximum d . Finally, u_t represents fluctuations from time-point to time-point imposed from outside the population, independent of density. It is easiest to understand this approach when the X_t s represent deviations from the long-term average abundance such that m disappears (the long-term average deviation from the mean is obviously zero). Then, in the absence of any density dependence (all β s zero) the abundance at time t will reflect simply the abundance at time $t - 1$ plus any ‘outside’ fluctuations u_t ; while any regulatory tendencies will be reflected in β values of less than zero.

Fennoscandian
microtines

Applying this approach to a time series of abundance (i.e. a sequence of X_t values) the usual first step is to determine the statistical model (X_t as the dependent variable) with the optimal number of time lags: the one that strikes the best balance between accounting for the variations in X_t and not including too many lags. Essentially, additional lags are included as long as they account for a significant additional element of the variation. The β values in the optimal model may then shed light on the manner in which abundance in the population is regulated and determined. An example is illustrated in Figure 14.9, which summarizes analyses of 19 time series of microtine rodents (lemmings and voles) from various latitudes in Fennoscandia (Finland, Sweden and Norway) sampled once per year (Bjørnstad *et al.*, 1995). In almost all cases, the optimum number of lags was two, and so the analysis proceeded on the basis of these two lags: (i) direct density dependence; and (ii) density dependence with a delay of 1 year.

Figure 14.9a sets out the predicted dynamics, in general, of populations governed by these two density dependences (Royama, 1992). Remember that delayed density dependence is reflected in a value of β_2 less than 0, while direct density dependence is reflected

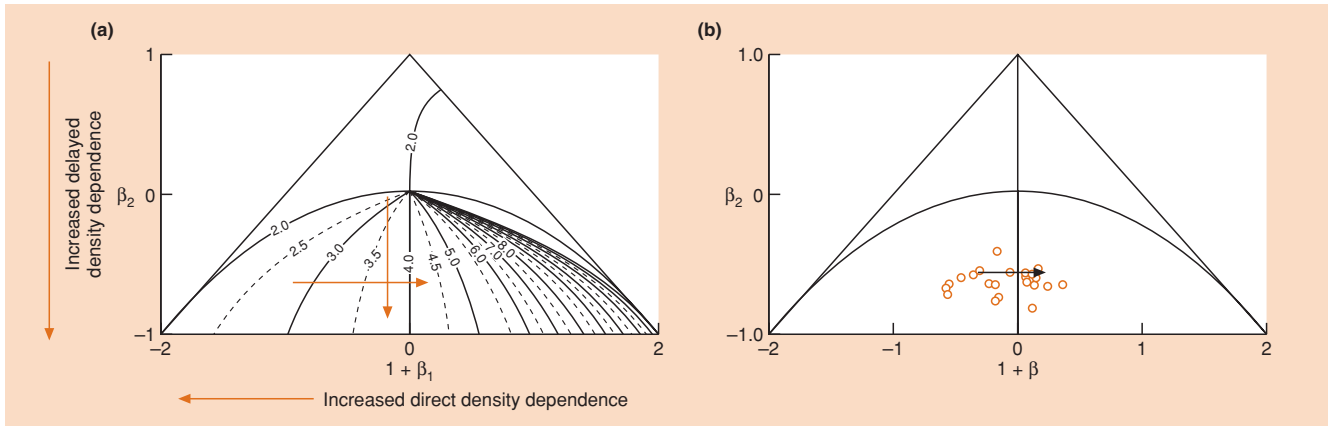


Figure 14.9 (a) The type of population dynamics generated by an autoregressive model (see Equation 14.1) incorporating direct density dependence, β_1 , and delayed density dependence, β_2 . Parameter values outside the triangle lead to population extinction. Within the triangle, the dynamics are either stable or cyclic and are always cyclic within the semicircle, with a period (length of cycle) as shown by the contour lines. Hence, as indicated by the arrows, the cycle period may increase as β_2 decreases (more intense delayed density dependence) and especially as β_1 increases (less intense direct density dependence). (b) The locations of the pairs of β_1 and β_2 values, estimated from 19 microtine rodent time series from Fennoscandia. The arrow indicates the trend of increasing latitude in the geographic origin of the time series, suggesting that a trend in cycle period with latitude, from around 3 to around 5 years, is the result of a decreased intensity of direct density dependence. (After Bjørnstad *et al.*, 1995.)

in a value of $(1 + \beta_1)$ less than 1. Thus, populations not subject to delayed density dependence tend not to exhibit cycles (Figure 14.9a), but β_2 values less than 0 generate cycles, the period (length) of which tends to increase both as delayed density dependence becomes more intense (down the vertical axis) and especially as direct density dependence becomes less intense (left to right on the horizontal axis).

support for the specialist predation hypothesis

The results of Bjørnstad *et al.*'s analysis are set out in Figure 14.9b. The estimated values of β_2 for the 19 time series showed no trend as latitude increased, but the β_1 values increased significantly. The points combining these pairs of β s, then, are shown in the figure, and the trend with increasing latitude is denoted by the arrow. It was known prior to the analysis, from the data themselves, that the rodents exhibited cycles in Fennoscandia and that the cycle length increased with latitude. The data in Figure 14.9b point to precisely the same patterns. But in addition, they suggest that the reasons lie in the structure of the density dependences: on the one hand, a strong delayed density dependence throughout the region, such as would result from the actions of specialist predators; and on the other hand, a significant decline with latitude in the intensity of direct density dependence, such as may result from an immediate shortage of food or the actions of generalist predators (see Figure 10.11b). As we shall see in Section 14.6.4 (see also Section 10.4.4), this in turn is supportive of the 'specialist predation' hypothesis for microtine cycles. The important point here, though, is the

illustration this example provides of the utility of such analyses, focusing on the abundances themselves, but suggesting underlying mechanisms.

14.5.2 Time series analysis: counting and characterizing lags

In other, related cases, the emphasis shifts to deriving the optimal statistical model because the number of lags in that model may provide clues as to how abundance is being determined. It may do so because Takens' theorem (see Section 5.8.5) indicates that a system that can be represented with three lags, for example, comprises three functional interacting elements, whereas two lags imply just two elements, and so on.

hares and lynx . . .

One example of this approach (another is described in Section 12.7.1) is the study of Stenseth *et al.* (1997) of the hare-lynx system in Canada to which we have already referred briefly in Section 10.2.5. We noted there that the optimal model for the hare time series suggested three lags, whereas that for the lynx suggested two. The density dependences for these lags are illustrated in Figure 14.10a. For the hares, direct density dependence was weakly negative (remember that the slope shown is $1 + \beta_1$) and density dependence with a delay of 1 year was negligible, but there was significant density dependence with a delay of 2 years. For the lynx, direct density dependence was effectively absent, but there was strong density dependence with a delay of 1 year.

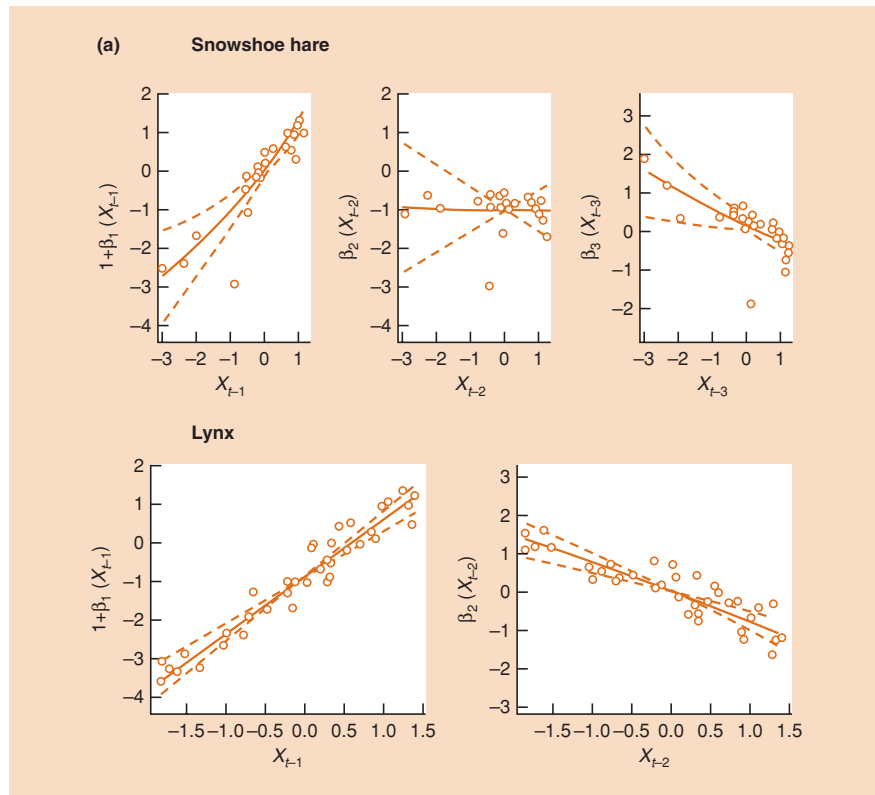


Figure 14.10 (a) Functions for the autoregressive equations (see Equation 14.1) for the snowshoe hare, above (three ‘dimensions’: direct density dependence and delays of 1 and 2 years), and the lynx, below (two dimensions: direct density dependence and a delay of 1 year). In each case, therefore, the slope indicates the estimated parameters, $1 + \beta_1$, β_2 and β_3 , respectively, reflecting the intensity of density dependence. The 95% confidence intervals are also shown.

... display three and two dimensions, respectively

This, combined with a detailed knowledge of the whole community of which the hare and lynx are part (Figure 14.10b, c), provided justification for Stenseth *et al.* (1997) to go on

to construct a three-equation model for the hares and a two-equation model for the lynx. Specifically, the model for the lynx comprised just the lynx and the hares, since the hares are by far the lynx’s most important prey (Figure 14.10b). Whereas the model for the hares comprised the hares themselves, ‘vegetation’ (since hares feed relatively indiscriminately on a wide range of vegetation), and ‘predators’ (since a wide range of predators feed on the hares and even prey on one another in the absence of hares, adding a strong element of self-regulation within the predator guild as a whole) (Figure 14.10c).

Lastly, then, and again without going into technical details, Stenseth *et al.* were able to recast the two- and three-equation models of the lynx and hare into the general, time-lag form of Equation 14.1. In so doing, they were also able to recast the β values in the time-lag equations as appropriate combinations of the interaction strengths between and within the hares, the lynx, and so on. Encouragingly, they found that these combinations were entirely consistent with the slopes (i.e. the β values) in Figure 14.10a. Thus, the elements that appeared to determine hare and lynx abundance were first counted (three and two, respectively) and

then characterized. What we have here, therefore, is a powerful hybrid of a statistical (time series) analysis of densities and a mechanistic approach (incorporating into mathematical models knowledge of the specific interactions impinging on the species concerned).

Finally, note that related methods of time series analysis have been used in the search for chaos in ecological systems, as described in Section 5.8.5. The motivations in the two cases, of course, are somewhat different. The search for chaos, none the less, is, in a sense, an attempt to identify as ‘regulated’ populations that appear, at first glance, to be anything but.

14.5.3 Combining density dependence and independence – weather and ecological interactions

Seeking to dissect out the relative contributions of direct and delayed density dependences, however, could itself be

multimammate rats in Tanzania

seen as prejudging the determinants of abundance by focusing too much on density-dependent, as opposed to density-independent, processes. Other studies, then, have examined time series precisely with a view to understanding how density-dependent and -independent factors combine in generating particular patterns

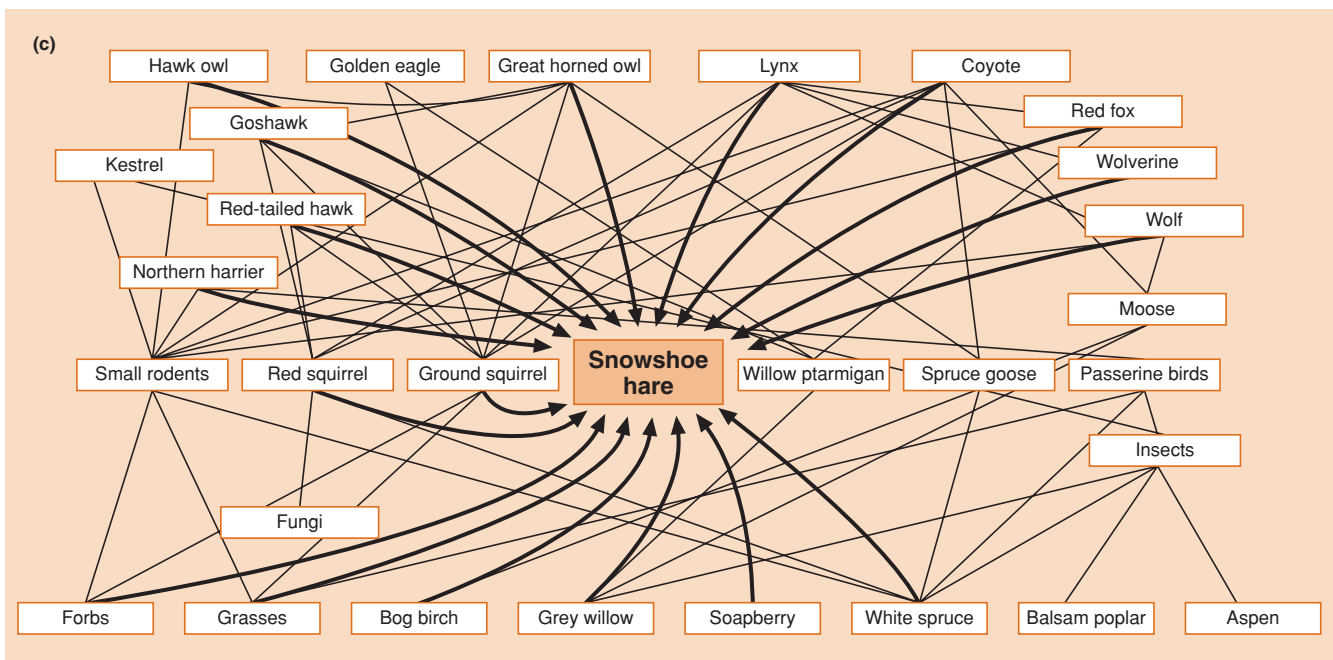
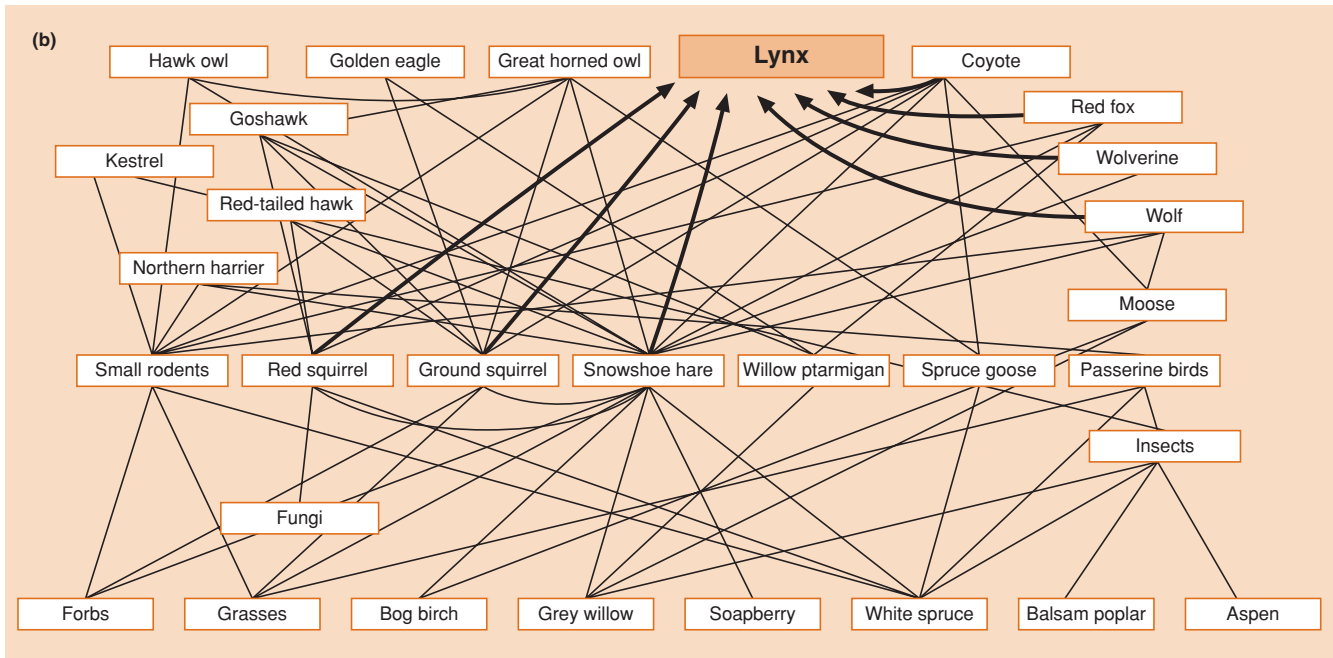


Figure 14.10 (continued) (b) The main species and groups of species in the boreal forest community of North America, with trophic interactions (who eats who) indicated by lines joining the species, and those affecting the lynx shown in bold. (c) The same community, but with the interactions of the snowshoe hare shown in bold. (After Stenseth *et al.*, 1997.)

of abundance. Leirs *et al.* (1997), for example, have examined the dynamics of the multimammate rat, *Mastomys natalensis*, in Tanzania. Using one part of their data to construct a predictive model (Figure 14.11a) and a second part to test that model's success (Figure 14.11b), they found first, in model construction, that variations in survival and maturation were far better accounted for by using both densities and preceding rainfall as predictors than by using either of these alone. In particular (Figure 14.11c), subadult survival probabilities showed no clear trends with either rainfall or density (though they tended to be higher at higher densities), but maturation rates increased markedly with rainfall (and were lowest at high densities following wet months), while adult survival was consistently lower at higher densities.

Estimates of demographic parameters (survival, maturation) from the statistical model were then used to construct a matrix model of the type described in Section 14.3.2, which was used in turn to predict abundance in the second, separate data set (Figure 14.11b), using rainfall and density there to predict 1 month ahead. The correspondence between the observed and predicted values was not perfect but was certainly encouraging

(Figure 14.11d). Hence, we can see here how the density, mechanistic (rainfall) and demographic approaches combine to provide insights into the determination of the rats' abundance. This example also reminds us that a proper understanding of abundance patterns is likely to require the incorporation of both density-dependent, biotic, deterministic effects and the density-independent, often stochastic effects of the weather.

Of course, not all the effects of the weather are wholly stochastic in the sense of being entirely unpredictable. mice and the ENSO in Chile

Apart from obvious seasonal variations, we saw in Section 2.4.1, for example, that there are a number of climatic patterns operating at large spatial scales and with at least some degree of temporal regularity, notably the El Niño–Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO). Lima *et al.* (1999) examined the dynamics of another rodent species, the leaf-eared mouse, *Phyllotis darwini*, in Chile and followed a similar path to Leirs and his colleagues in combining the effects of ENSO-driven rainfall variability and delayed density dependence in accounting for the observed abundance patterns.

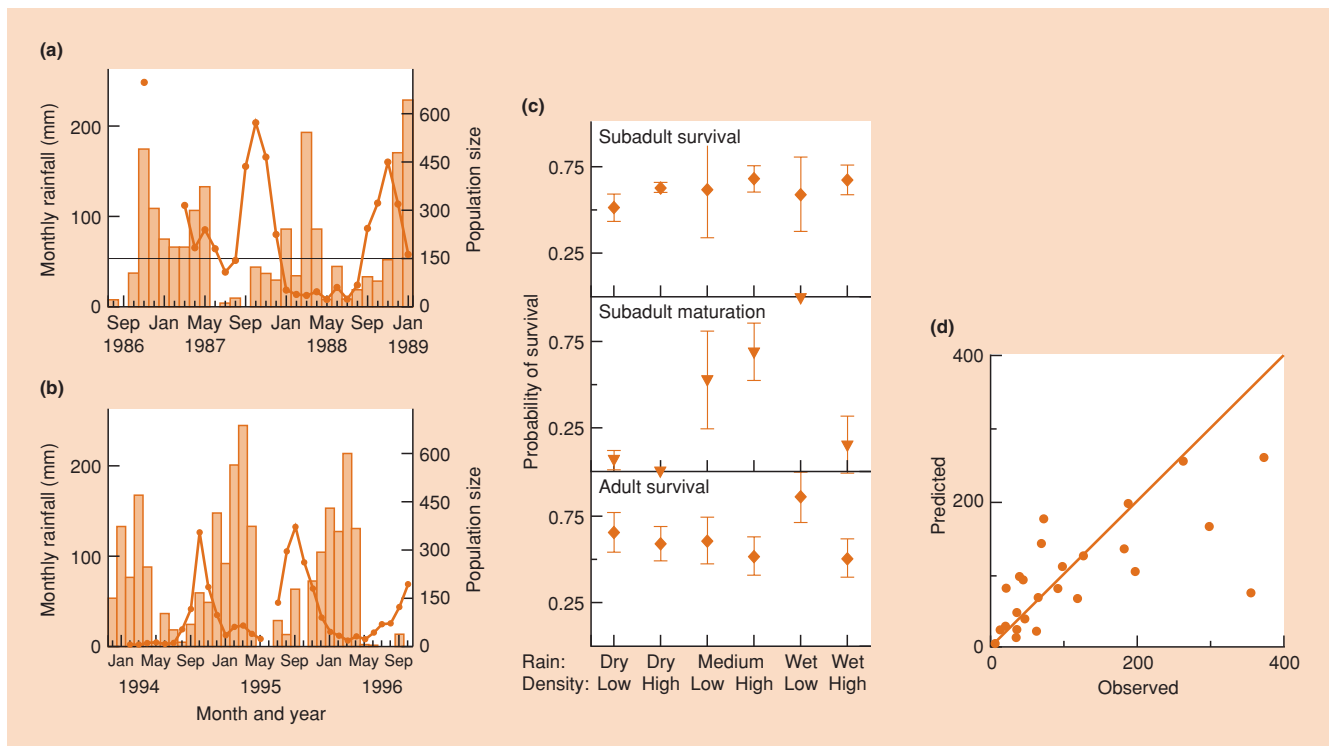


Figure 14.11 (a) Time series data for the multimammate rat (dots) and rainfall (bars) in Tanzania, used to derive a statistical model to predict rat abundance. (The horizontal line indicates the cut-off between 'high' and 'low' densities.) (b) Subsequent time series data used to test that model. (c) Estimates (\pm SE) from the model for the effects of density and rainfall on population size. (d) The relationship between predicted and observed population sizes in the test data ($r^2 = 0.49$, $P < 0.001$); the line of equality is also shown. (After Leirs *et al.*, 1997.)

14.6 Population cycles and their analysis

Regular cycles in animal abundance were first observed in the long-term records of fur-trading companies and gamekeepers. Cycles have also been reported from many studies of voles and lemmings and in certain forest Lepidoptera (Myers, 1988). Population ecologists have been fascinated by cycles at least since Elton drew attention to them in 1924. In part, this fascination is attributable to the striking nature of a phenomenon that is crying out for an explanation. But there are also sound scientific reasons for the preoccupation. First, cyclic populations, almost by definition, exist at different times at a wide range of densities. They therefore offer good opportunities (high statistical power) for detecting such density-dependent effects as might exist, and integrating these with density-independent effects in an overall analysis of abundance. Furthermore, regular cycles constitute a pattern with a relatively high ratio of 'signal' to 'noise' (compared, say, to totally erratic fluctuations, which may appear to be mostly noise). Since any analysis of abundance is likely to seek ecological explanations for the signal and attribute noise to stochastic perturbations, it is obviously helpful to know clearly which is signal and which is noise.

extrinsic and intrinsic factors

Explanations for cycles are usually classified as emphasizing either extrinsic or intrinsic factors. The former, acting from outside the population, may be food, predators or parasites, or some periodic fluctuation in the environment itself. Intrinsic factors are changes in the phenotypes of the organisms themselves (which might in turn reflect changes in genotype): changes in aggressiveness, in the propensity to disperse, in reproductive output, and so on. Below we examine studies on population cycles in three systems, all of which we have touched on previously: the red grouse (Section 14.6.2), the snowshoe hare (and lynx) (Section 14.6.3) and microtine rodents (Section 14.6.4). In each case, it will be important to bear in mind the problems of disentangling cause from effect; that is, of distinguishing factors that change density from those that merely vary *with* density. Equally, it will be important to try to distinguish the factors that affect density (albeit in a cyclic population) from those that actually impose a pattern of cycles (see also Berryman, 2002; Turchin, 2003).

14.6.1 Detecting cycles

The defining feature of a population cycle or oscillation is regularity: a peak (or trough) every x years. (Of course, x varies from case to case, and a certain degree of variation around x is inevitable; even in a '3-year cycle', the occasional interval of 2 or 4 years is to be expected.) The statistical methods applied to a time series, to determine whether the claim of 'cyclicity' can justifiably be made, usually involve the use of an autocorrelation

function (Royama, 1992; Turchin & Hanski, 2001). This sets out the correlations between pairs of abundances one time interval apart, two time intervals apart, and so on (Figure 14.12a). The correlation between abundances just one time interval apart can often be high simply because one abundance has led directly to the next. Thereafter, a high positive correlation between pairs, for example, 4 years apart would indicate a regular cycle with a period of 4 years; while a further high *negative* correlation between pairs 2 years apart would indicate a degree of symmetry in the cycle: peaks and troughs typically 4 years apart; with peaks typically 2 years from troughs.

It must be remembered, however, that it is not just the pattern of an autocorrelation function that is important but also its statistical significance.

autocorrelation function analysis

Even a single clear rise and fall in a relatively short time series may hint at a cycle (Figure 14.12b), but this pattern would need to be repeated in a much longer series before the autocorrelations were significant, and only then could a cycle be said to have been identified (and require explanation). It is no surprise that major investments in time and effort are required to study cycles in natural populations. Even where those investments have been made, the resulting 'ecological' time series are shorter than those commonly generated in, say, physics – and shorter than those probably envisaged by the statisticians who devised methods for analyzing them. Ecologists need always to be cautious in their interpretations.

14.6.2 Red grouse

The explanation for cycles in the dynamics of the red grouse (*Lagopus lagopus scoticus*) in the United Kingdom has been a matter of disagreement for decades. Some have emphasized an extrinsic factor, the parasitic nematode *Trichostrongylus tenuis* (Dobson & Hudson, 1992; Hudson *et al.*, 1998). Others have emphasized an intrinsic process through which increased density leads to more interactions between non-kin male birds and hence to more aggressive interactions. This leads in turn to wider territorial spacing and, with a delay because this is maintained into the next year, to reduced recruitment (Watson & Moss, 1980; Moss & Watson, 2001). Both viewpoints, therefore, rely on a delayed density dependence to generate the cyclic dynamics (see Section 10.2.2), though these are arrived at by very different means.

We have already seen in Sections 12.7.4 and 14.4.2 that even field-scale experiments have been unable to determine the role of the nematodes with certainty.

parasites?

There seems little doubt that they reduce density, and the results of the experiment are consistent with them generating the cycles, too. But the results are also consistent with the nematodes determining the amplitude of the cycles but not generating them in the first place.

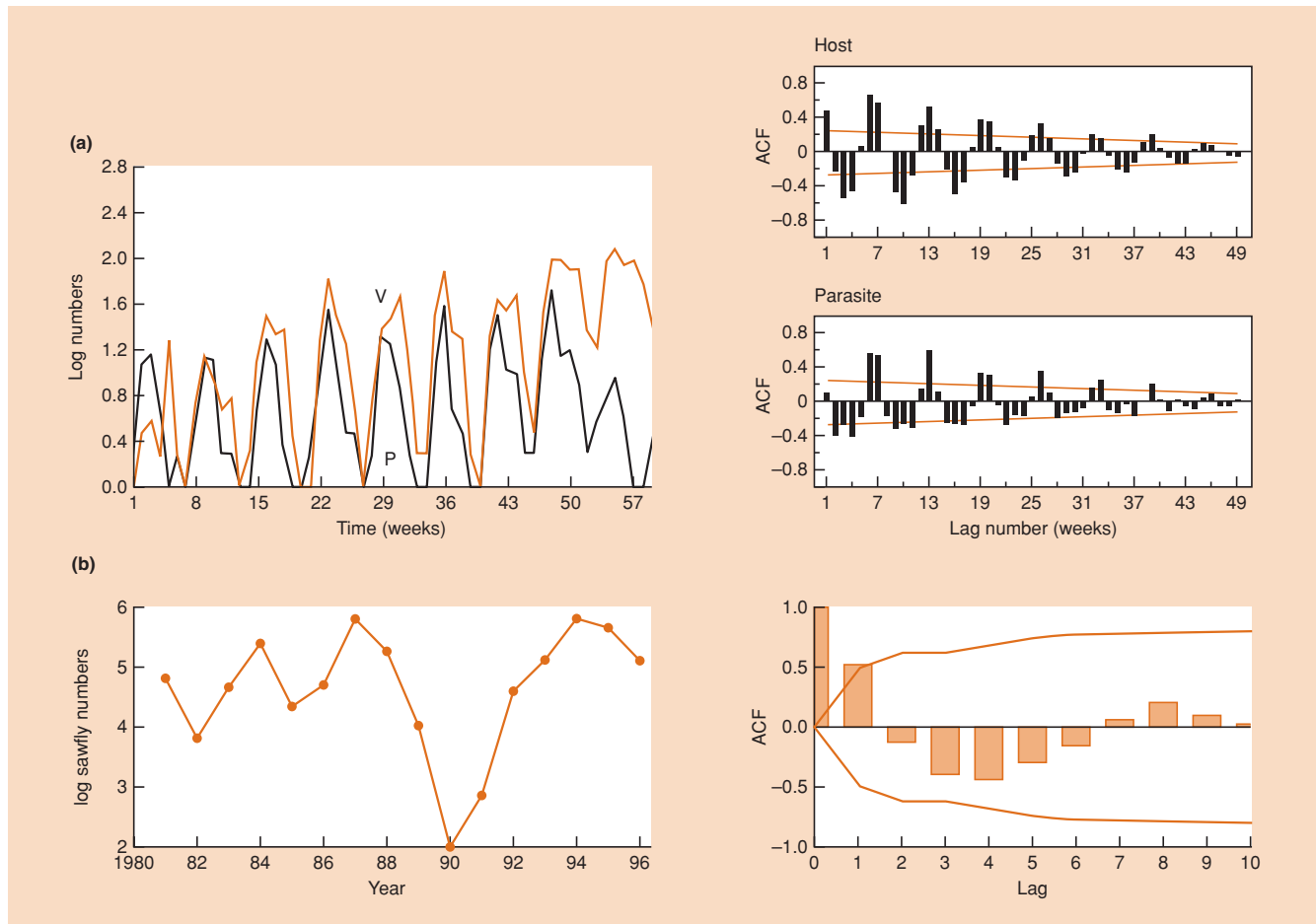


Figure 14.12 (a) Coupled oscillations in the abundance of the moth *Plodia interpunctella* and its parasitoid *Venturia canescens* (P and V, respectively) and, on the right, an autocorrelation function (ACF) analysis of those data (host above, parasitoid below). Sloping lines show the levels the bars must exceed for statistical significance ($P < 0.05$). The cycle periods (l) are 6–7 weeks, with significant correlations at l , $2l$, etc. and significant negative correlations at $0.5l$, $1.5l$, etc. (After Begon *et al.*, 1996.) (b) Time series for the abundance of the sawfly *Euura lasiolepis* (left) and an ACF analysis of those data (right). There is a hint of an 8-year cycle (positive correlation with a lag of 8 years; negative correlation with a lag of 4 years), but this does not come close to statistical significance (exceeding the lines). (After Turchin & Berryman, 2000; following Hunter & Price, 1998).

kinship, territories and aggression?

In another field experiment, aspects of the alternative, ‘kinship’ or ‘territorial behavior’ hypothesis were tested (Mougeot *et al.*, 2003). In experimental areas, established males were given testosterone implants at the beginning of the fall, when territorial contests take place. This increased their aggressiveness (and hence the size of their territories) at densities that would not normally generate such aggression. By the end of the fall, it was clear that, relative to the control areas, the increased aggression of the older males had reduced the recruitment of the younger males: testosterone treatment had significantly reduced male densities and had particularly reduced

the ratio of young (newly recruited) to old males, though there was no consistent effect on female densities (Figure 14.13a).

Moreover, in the following year, even though the direct effects of the testosterone had worn off, the young males had not returned (Figure 14.13a). Also, because young *relatives* had been driven out, levels of kinship were likely to be lower in experimental than in control areas. Hence, the kinship hypothesis predicts that recruitment and density in the experimental areas should have remained lower through the following year: that is, lower kinship leads to more aggression, which leads to larger territories, which leads to lower recruitment, which leads to lower density. These predictions were borne out (Figure 14.13b).

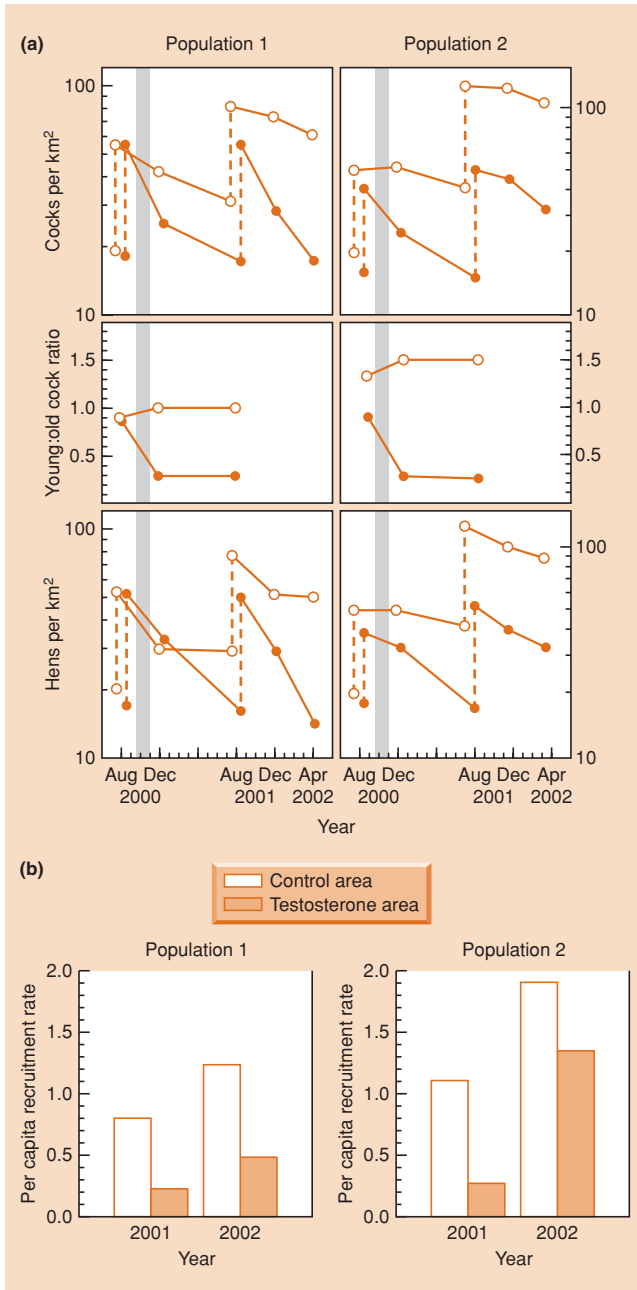


Figure 14.13 (a) Changes in grouse numbers (males (cocks); the young : old cock ratio; and females (hens)) in control (○) and testosterone-implant experimental areas (●) in two populations. The gray bar represents the period of time over which the males were given implants. (b) Per capita recruitment in the two populations was higher in the control areas than in the experimental areas, both in 2001, immediately after treatment, and 1 year later. (After Mougeot *et al.*, 2003.)

Thus, these results establish, at least, the potential for intrinsic processes to have (delayed) density-dependent effects on recruitment, and thus to generate cycles in the grouse. In a companion paper, Matthiopoulos *et al.* (2003) demonstrate how changes in aggressiveness can cause population cycles. As Mougeot *et al.* themselves note, though, the possibility remains that both the parasites and territorial behavior contribute to the observed cycles. Indeed, the two processes may interact: parasites, for example, reduce territorial behavior (Fox & Hudson, 2001). Certainly, it is far from guaranteed that either of the alternative explanations will ultimately be declared the ‘winner’.

14.6.3 Snowshoe hares

The ‘10-year’ hare and lynx cycles have also been examined in previous sections. We have seen, for example, from Stenseth *et al.*’s (1997) analysis of time series (see Section 14.5.2) that despite becoming a ‘textbook’ example of coupled predator–prey oscillations, the hare cycle appears in fact to be generated by interactions with both its food and its predators, both considered as guilds rather than as single species. The lynx cycle, on the other hand, does indeed appear to be generated by its interaction with the snowshoe hare.

This supports other results obtained by much more direct, experimental means, reviewed by Krebs *et al.* (2001). The demographic patterns underlying the hare cycle are relatively clear cut: both fecundity and survival begin to decline well before peak densities are reached, arriving at their minima around 2 years after density has started to decline (Figure 14.14).

First, we can ask: ‘What role does the hares’ interaction with their food play in these patterns?’ A whole series of field experiments in which artificial food was added, or natural food was supplemented, or food quality was manipulated either by fertilizers or by cutting down trees to make high-quality twigs available, all pointed in the same direction. Food supplementation may improve individual condition and in some cases lead to higher densities, but food by itself seems to have no discernible influence on the cyclic pattern (Krebs *et al.*, 2001).

field-scale
manipulations
of food and/or
predators

On the other hand, experiments in which either predators were excluded, or they were excluded and food was also supplemented, had much more dramatic effects. In the study by Krebs *et al.* (1995) carried out at Kluane Lake in the Yukon, Canada (Figure 14.15a), the combination of the two treatments all but eliminated the pattern of decline in survival over the cycle from 1988 to 1996, and predation played by far the major role in this.

Furthermore, food supplementation reduced slightly the initial decline in fecundity prior to peak densities (Figure 14.15b), but the combination of food supplementation and predator exclusion brought fecundity up to almost maximum levels at the phase of

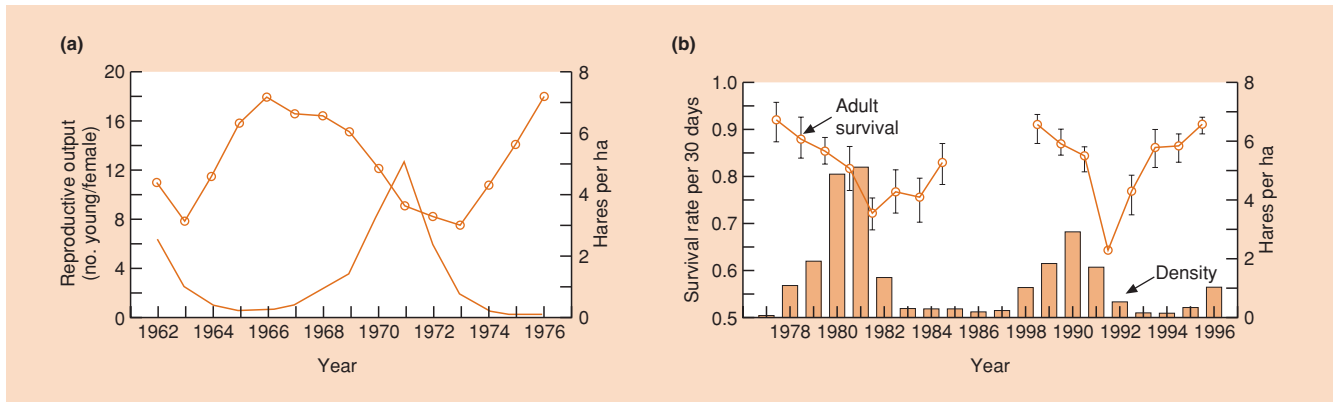


Figure 14.14 (a) Variation in reproductive output per year (dots) as density (continuous line) changes over a snowshoe hare cycle in central Alberta, Canada. (b) Variation in survival over two snowshoe hare cycles at Kluane Lake, Yukon, Canada. Too few hares were caught to estimate survival between 1985 and 1987. (After Krebs *et al.*, 2001; (a) following Cary & Keith, 1979.)

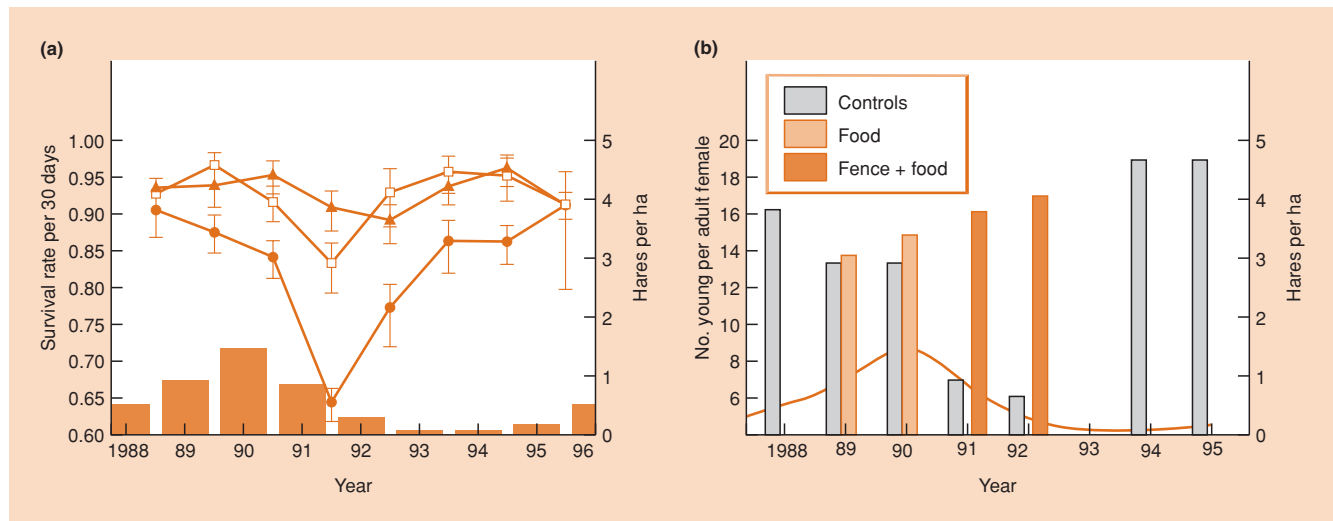


Figure 14.15 (a) Survival of radio-collared hares (with 90% confidence intervals) over a hare cycle from 1988 to 1996 at Kluane Lake, Yukon, Canada. The bars are densities; lines show the survival in controls (●) with mammalian predators excluded (□) and with mammalian predators excluded and food supplemented (▲). (b) Reproductive output over a hare cycle from 1988 to 1995 at Kluane Lake (the line). It was possible to compare control values with those from treatments of food supplementation in 1989 and 1990, and with those where food was supplemented and mammalian predators excluded in 1991 and 1992. (After Krebs *et al.*, 2001; (a) following Krebs *et al.*, 1995.)

lowest fecundity following the density peak. Unfortunately, it was not possible to measure fecundity in a treatment where only food was supplemented – an example of the disappointments that almost inevitably accompany large field experiments – so the effects of food and predators cannot be disentangled. Of these, any effects of food shortage on fecundity would be easy to understand. It is also possible, though, that an increased frequency of interaction with predators could reduce fecundity through its physiological effects on hares (reduced energy or increased levels of stress-associated hormones).

Thus, these hard-won results from sunspot cycles? field experiments and the analyses of time series essentially agree in suggesting that the snowshoe hare cycle results from interactions with both its food and its predators, with the latter playing the dominant role. It is also noteworthy that, at least over some periods, there has been a high correlation between the hare cycle and the 10-year cycle of sunspot activity, which is known to affect broad weather patterns (Sinclair & Gosline, 1997). This type of extrinsic, abiotic factor was initially a strong candidate for playing a major role in driving population

cycles generally (Elton, 1924). Subsequently, however, they have received little support. In the first place, many population cycles are of the wrong period and are also variable in period (see, for example, the microtine rodents in the next section). Second, the population cycles are often more pronounced than the extrinsic cycles that are proposed to be ‘causing’ them. Also, even when a correlation has been established, as in the present case, this simply begs the question of what links the two cycles: presumably it is climate acting on some combination of the factors we have already been considering – predators, food and intrinsic features of the population itself – although no mechanistic basis for such a link has been established.

Overall, then, the snowshoe hare work illustrates how a range of methodologies may come together in the search for an explanation of a cyclic pattern. It also provides a very sobering reminder of the logistical and practical difficulties – collecting long time series, undertaking large field experiments – that need to be accepted and overcome in order to build such explanations.

14.6.4 Microtine rodents: lemmings and voles

many microtines cycle – and many don’t

There is no doubt that more effort has been expended overall in studying population cycles in microtine rodents (voles and lemmings) than in any other group of species. Cycle periods are typically 3 or 4 years, or much more rarely 2 or 5 years or even longer. These cyclic dynamics have been convincingly identified in a range of communities, including the following: voles (*Microtus* spp. and *Clethrionomys* spp.) in Fennoscandia (Finland, Norway and Sweden); lemmings (*Lemmus lemmus*) elsewhere in montane habitats in Fennoscandia; lemmings (*Lemmus* spp. and *Dicrostonyx* spp.) in the tundra of North America, Greenland and Siberia; voles (*Clethrionomys rufocanus*) in Hokkaido, northern Japan; common voles (*Microtus arvalis*) in central Europe; and field voles (*Microtus agrestis*) in northern England. On the other hand, there are many other microtine populations that show no evidence of multiannual cycles, including voles in southern Fennoscandia, southern England, elsewhere in Europe, and many locations in North America (Turchin & Hanski, 2001). It is also worth emphasizing that a quite different pattern, of irregular and spectacular irruptions in abundance and mass movement, is shown by just a few lemming populations, notably in Finnish Lapland. It is these whose suicidal behavior has been so grossly exaggerated (to say the least) in the name of film-makers’ poetic license, unfairly condemning all lemmings to popular misconception (Henttonen & Kaikusalo, 1993).

trends in cyclicity

Over many decades, the same range of extrinsic and intrinsic factors have been proposed to explain microtine cycles as have been directed at population cycles generally. Given the variety of species and habitats, it is perhaps especially

unlikely in this case that there is a single all-encompassing explanation for all of the cycles. None the less, there are a number of features of the cycles that any explanation, or suite of explanations, must account for. First is the simple observation that some populations cycle while others do not. Also, there are cases (notably in Fennoscandia) where several coexisting species, often with apparently quite different ecologies, all cycle synchronously. And there are sometimes clear trends in cycle period, notably with increasing latitude (south to north) in Fennoscandia (see Section 14.5.1), where an explanation has been most intensively sought, but also for example in Hokkaido, Japan, where cyclicity increases broadly from southwest to northeast (Stenseth *et al.*, 1996), and in central Europe, where cyclicity increases from north to south (Tkadlec & Stenseth, 2001).

A useful perspective from which to proceed is to acknowledge, as we have seen, that the rodent cycles are the result of a ‘second-order’ process (Bjørnstad *et al.*, 1995; Turchin & Hanski, 2001) (see Section 14.5.1); that is, they reflect the combined effects of a directly density-dependent process and a delayed density-dependent process. This immediately alerts us to the fact that, in principle at least, the direct and delayed processes need not be the same in every cyclic population: what is important is that two such processes act in conjunction.

We start with the ‘intrinsic’ theories. It is not surprising that voles and lemmings, which can achieve extremely high potential rates of population growth, should experience periods of overcrowding. Neither would it be surprising if overcrowding then produced changes in physiology and behavior. Mutual aggression (even fighting) might become more common and have consequences in the physiology, especially the hormonal balance, of the individuals. Individuals may grow larger, or mature later, under different circumstances. There might be increased pressure on some individuals to defend territories and on others to escape. Kin and non-kin might behave differently to one another when they are crowded. Powerful local forces of natural selection might be generated that favor particular genotypes (e.g. aggressors or escapists). These are responses that we easily recognize in crowded human societies, and ecologists have looked for the same phenomena when they try to explain the population behavior of rodents. All these effects have been found or claimed by rodent ecologists (e.g. Lidicker, 1975; Krebs, 1978; Gaines *et al.*, 1979; Christian, 1980). But it remains an open question whether any of them plays a critical role in explaining the behavior of rodent populations in nature.

In the first place, we saw in Sections 6.6 and 6.7 the complexities of the relationships in rodents between density, dispersal, relatedness and ultimate survival and reproductive success. What is more, by no means all of this work has been carried out on species that exhibit cycles. Hence, there is little support for any universal rules, but there do seem to be tendencies

cycles result from a second-order process

dispersal, relatedness and aggression?

for most dispersal to be natal (soon after birth), for males to disperse more than females, for effective dispersal (arriving, rather than simply traveling hopefully) to be more likely at lower densities, and for fitness to be greater the greater the relatedness of neighbors. This has led some in the field to argue that the ‘jury is still out’ (Krebs, 2003), but others have simply doubted any role for these processes in the regulation of rodent populations, especially in view of the frequent *inverse* density dependence (Wolff, 2003). Certainly, while variations between individuals may be associated with different phases of the cycle, this is very far from saying that they are *driving* the cycles. If individuals disperse more at particular cycle phases, say, or are larger, then this is likely to be a *response* either to a present or to a past level of food or space availability, or to predation pressure or infection intensity. That is, intrinsic variations are more likely to explain the detailed nature of responses, whereas extrinsic factors are more likely to explain the causes of the responses.

maternal effects?

None the less, in one case at least, an intrinsic cause has been proposed for the delayed density dependence. Inchausti and Ginzburg (1998) constructed a model with a ‘maternal effect’, in which mothers transmit their body condition phenotypically to their daughters, either from spring to fall or from fall to spring, and this in turn determines their per capita rate of growth. Thus, in this case the intrinsic quality of an individual is indeed a response to a past density, and hence to past resource availability, accounting for the delayed density dependence. Furthermore, when Inchausti and Ginzburg, focusing on Fennoscandia, fed what they believed to be reasonable values of population growth rate and the maternal effect into their model, both decreasing with latitude, they were able to recreate cycles with periods varying from 3 to 5 years (Figure 14.16). Turchin and Hanski (2001) criticized the parameter estimates (especially those of the growth rates) and claimed that the maternal effect model actually predicted 2-year cycles, at odds with those observed. Ergon *et al.* (2001) found with field voles, *Microtus agrestis*, from cyclic populations, that in transferring them between contrasting sites they rapidly took on characteristics appropriate to their new rather than their old populations – and certainly not those of their mothers. None the less, Inchausti and Ginzburg’s results, set alongside the specialist predation hypothesis (see Section 14.5.1 and below), emphasize how the same pattern (here, the latitudinal gradient) might be achieved by quite different means. They also show that intrinsic theories remain ‘in play’ in the continuing search for an explanation for microtine cycles.

the specialist predation hypothesis

Turning now to extrinsic factors, there are two main candidates: predators and food. (Parasites and pathogens excited Elton’s interest immediately after his original, 1924 paper, but they were largely ignored subsequently until recent technical advances made their study a serious possibility. It remains to be seen what role, if any, they

None the less, in one case at least, an intrinsic cause has been proposed for the delayed density dependence.

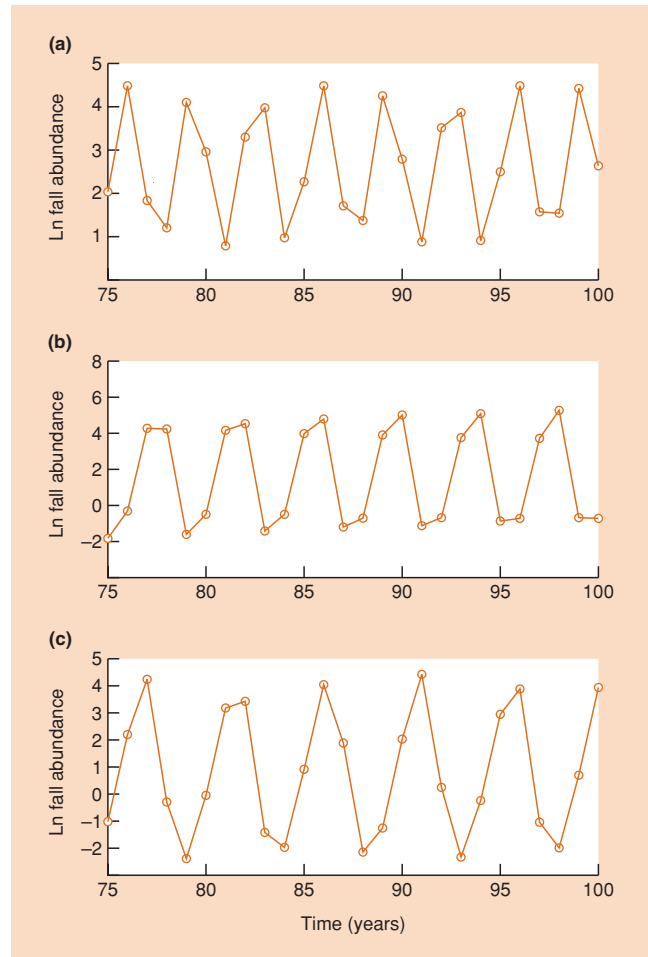


Figure 14.16 Behavior of Inchausti and Ginzburg’s (1998) maternal effect model with differing values of the maximum yearly reproductive rate, R , and the maternal effect, M , through which the quality of daughters in one season is affected by the quality of mothers in the previous season (fall or spring). The simulations are given 75 ‘years’ to settle into a regular pattern. (a) $R = 7.3$; $M = 15$. (b) $R = 4.4$; $M = 10$. (c) $R = 3.5$; $M = 5$. (After Inchausti & Ginzburg, 1998.)

play.) We have already made a start in examining predators in Sections 10.4.4 and 14.5.1. Their importance in microtine cycles, expressed as the ‘specialist predation hypothesis’, has received considerable support since around 1990 from a series of mathematical models and field experiments, especially from workers focused on the cycles in Fennoscandia. The hypothesis, put simply, is that specialist predators are responsible for the delayed density dependence, whereas generalist predators, whose importance varies with latitude, are a major source of direct density dependence.

Early field experiments in which predators were removed (in Fennoscandia and elsewhere), although they

experimental support?

typically led to 2–3-fold increases in vole density, were subject to various criticisms of their experimental design: they were short term, or small scale, or they affected too many, or too few, of the predator species, and they often involved the erection of protective fences that are likely to have affected movements of the prey (voles) as well (Hanski *et al.*, 2001). Conclusive experiments may be a necessity, but this does not make them any easier! More recent experiments, too, give rise to some of the same misgivings. Klemola *et al.* (2000) excluded all the predators from four fenced (and net-roofed) exclosures in western Finland, 1 ha in size, for 2 years. The vole populations in the exclosures increased more than 20-fold in abundance compared to the control grids, until food shortages caused them to crash (Figure 14.17a). But the effects of specialists and generalists were inevitably combined in such a design; and while results such as these indicate an important role for predators in vole survival and abundance, they cannot prove

a role in causing (as opposed, say, to amplifying) the vole cycles. Korpimäki *et al.* (2002), worked in the same area but used four much larger unfenced areas (2.5–3 km²) for 3 years, reducing predator abundance over the summer but not the winter: mustelids (stoats and weasels) by trapping, and avian predators by removing natural and artificial nesting sites. Predator reduction increased vole density fourfold in the first (low) year; it accelerated an increase in density twofold in the second year; and it increased fall density twofold in the third (peak) year (Figure 14.17b). But again, specialists and generalists were not distinguished, and the temporal pattern of abundance was essentially unaltered.

The specialist predation model itself, which has been successively refined in a series of studies (the refinements are traced by Hanski *et al.*, 2001) has the following key features: (i) logistic population growth in the microtine prey, to reflect the directly density-dependent effects of food shortage on the microtines, preventing their populations from growing too large before specialist predators ‘catch up’; (ii) specialist predators (weasels) with a population growth rate that declines as the ratio of specialist predators to prey increases; (iii) seasonal differences in the breeding of voles and weasels in the summer and winter; and (iv) generalist predators – generalist, switching mammals or wide ranging (nomadic) avian specialists that act in a directly density-dependent manner by responding immediately to changes in microtine density. Note, therefore, that the model includes both of the most-studied extrinsic factors: predators *and* food. Food provides the baseline direct density dependence; specialist predators provide the delayed density dependence. The generalist predators then provide a further source of direct density dependence that can be varied to mimic their known decline in abundance with latitude.

When the model is parameterized with field data from Fennoscandia, it can recreate an impressive number of the features of the observed dynamics. Cycles are of broadly the correct amplitude and period, and both the period and indeed the amplitude of the cycles increase with latitude as the density of generalist predators decreases, as observed in nature (Figure 14.18). A related model for the collared lemming, *Dicrostonyx groenlandicus*, preyed upon by one specialist predator (the stoat, *Mustela erminea*) and three generalists (Gilg *et al.*, 2003) was also able to recreate observed cycles in Greenland when parameterized with field data.

On the other hand, not all studies have conformed to the predictions of the specialist predation model. Lambin *et al.* (2000) described regular cycles of field voles in Kielder forest, northern England (55°N), with a period of 3–4 years and an approximately 10-fold difference between peak and trough densities (a difference of 1 on a log scale, such as on Figure 14.18). Yet, parameterizing the specialist predation model with the estimated intensity of generalist predation at this site would have predicted no cycles whatsoever – as would the latitude. What is more, a rigorous

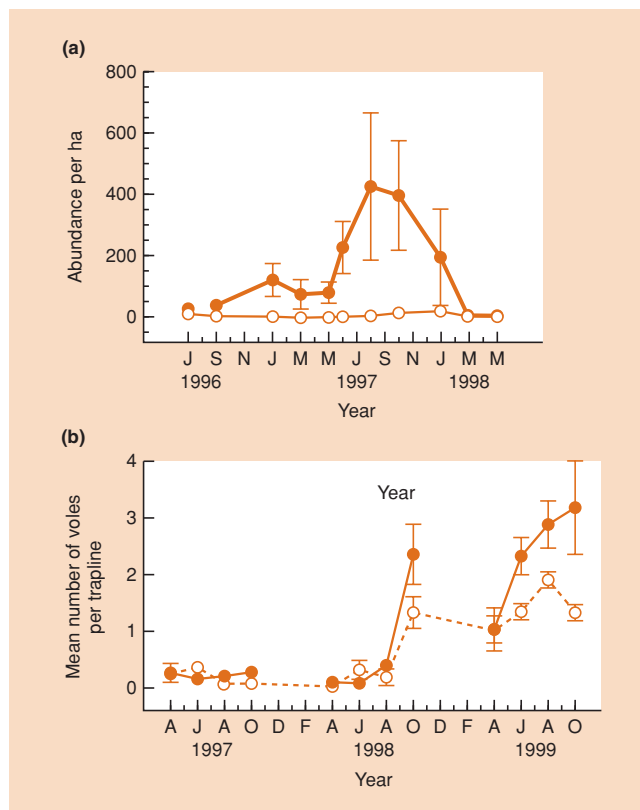


Figure 14.17 (a) Mean abundances of voles (\pm SE) from four small predator-exclosure grids (●) and four control grids (○) in western Finland. (After Klemola *et al.*, 2000.) (b) The density of voles (mean number of individuals caught per trapline, \pm SE, in April, June, August and October) from four large predator-reduction sites (●) and four control sites (○) in western Finland. Predator reduction occurred only over the summer and vole densities tended to revert to control levels over the winter. (After Korpimäki *et al.*, 2002.)

supportive
prediction?

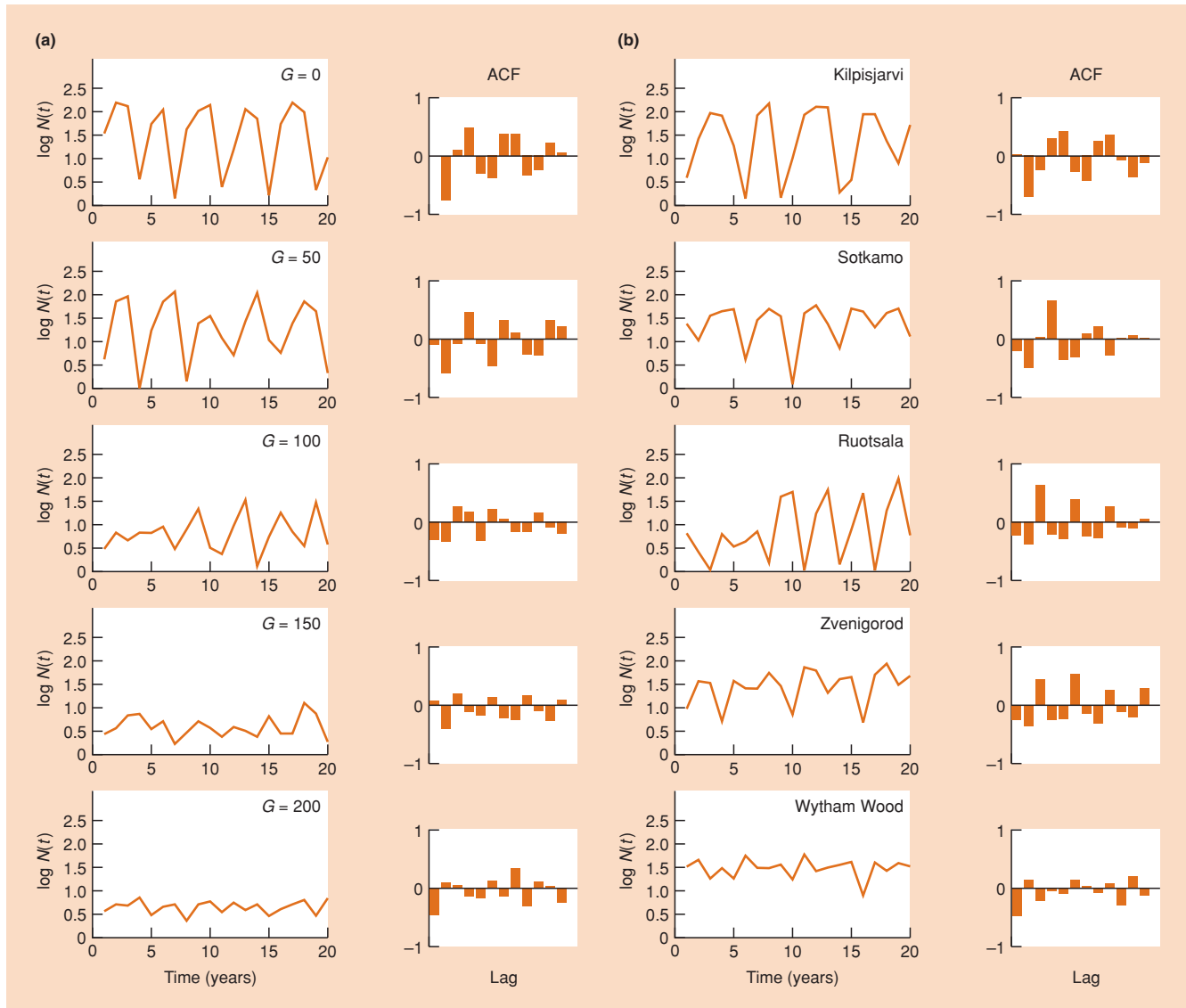


Figure 14.18 (a) Sample data generated by the specialist predation model, and the associated autocorrelation functions (ACFs), for various values of generalist predator abundance, G . As G increases, cycle period increases and cycle amplitude decreases, and at sufficiently high values the dynamics are sufficiently stabilized for the cycles to disappear altogether. (b) Comparable time series from five field sites: Kilpisjarvi (69°N ; period = 5), Sotkamo (64°N ; period = 4), Ruotsala (63°N ; period = 3), Zvenigorod (57°N ; period = 3) and Wytham Wood (51°N ; no significant periodicity). (After Turchin & Hanski, 1997.)

program of reducing weasel numbers (i.e. specialist predators) at unfenced grids within the site (by about 60% in comparison with control sites) increased adult vole survival by about 25% but had no appreciable impact on the cyclic dynamics (Graham & Lambin, 2002).

Lambin and his collaborators conclude from these studies that generalist predators may not, after all, be responsible for the Fennoscandian gradient in cycle length; and that vole cycles need not be the result of the impact of specialist (i.e. weasel)

predation (since they seem not to be at Kielder). Remember, too, that the results of time series analyses (see Section 14.5.1) and predator removal studies in Fennoscandia are consistent with the specialist predation hypothesis but do not prove it. In contrast, the response to these results of adherents of the specialist predation hypothesis (e.g. Korpimäki *et al.*, 2003) has been to emphasize that the Kielder cycles are different from those in northern Fennoscandia (lower amplitude at Kielder, higher trough densities, less spatial synchrony and only one vole species

involved). That is, they contend that the results at Kielder may tell us little or nothing about cycles in Fennoscandia. Though we may strive to avoid it, even rigorous studies are often open to alternative interpretations.

a role for food?

Turning finally to the role of food, both field observations and experiments suggest that it would be unwise to assume that the same forces act on voles and lemmings (Turchin & Batzli, 2001). In the first place, voles typically eat a range of vascular plants, including graminoids (grasses and sedges), whereas lemmings feed on a mixture of mosses and graminoids. Voles seem rarely to consume more than a few percent of available plant material (though, of course, the quality of the available food might be more important than its quantity – see, for example, Batzli, 1983); and food supplementation has typically failed to increase vole abundance (though experiments may have been foiled by the ‘pantry effect’ through which predators are attracted to high vole densities, counteracting the effects of supplementation). Lemmings, on the other hand, at peak densities, typically remove more than 50% and sometimes as much as 90–100% of the available vegetation.

Furthermore, through an analysis of models, Turchin and Batzli (2001) show clearly that the role that vegetation might play in cyclic dynamics depends critically on the nature of the vegetation itself, especially the vegetation’s dynamics following significant consumption by herbivores. If the dynamics are logistic (i.e. S-shaped) then this may provide the delayed density dependence necessary to generate ‘second-order’ cycles in microtine abundance. But if the dynamics are ‘regrowth’ (i.e. a rapid initial response, decelerating until a saturation abundance is reached) then any density dependence will be direct rather than delayed. In this case, the microtine–food interaction may play an integral part in cyclic dynamics (as they do, for example, in the specialist predation hypothesis), but it cannot be the second-order driving force. Crucially, plants consumed by voles seem likely to exhibit rapid regrowth dynamics because of the large proportion of unconsumed plant parts, much of it underground. By contrast, mosses are, by their nature, wholly available to their consumers, and when lemmings devastate their vegetation they often grub underground for graminoid rhizomes and destroy these too. Lemming vegetation, therefore, is likely to exhibit logistic dynamics: rapid only after a slow start.

On this basis, Turchin and Batzli parameterized a model for microtines and a food supply with logistic growth, using what data were available for the brown lemming (*Lemmus sibiricus*) and its vegetation at Barrow, Alaska (Batzli, 1993). The results were encouraging although not perfect representations of the observed patterns: cycle amplitudes were too low (400- rather than 600-fold) and too long (6 rather than 4 years). On the other hand, uncertainty, and in some cases plain ignorance, surrounded several of the parameter estimates. The model can be ‘tweaked’ to generate the observed dynamics. Further painstaking work in the field,

especially to obtain winter parameter estimates from under the snow, will be required to determine whether such tweaking is justified by the truth about lemming biology.

The cycles of microtines have been studied for longer and with greater intensity than those of any other species, and have generated more theories to explain them, and more disagreements amongst disputing advocates. At the time of writing, a near-consensus appears to have been arrived at that a conjunction of direct and delayed density dependence is required to account for observed patterns; and most support is attracted to the contention that specialist predators provide the delayed density dependence, while food shortage and generalist predators provide the direct density dependence. All scientific ‘conclusions’ are provisional, however, and fashions change in science as in everything else. It remains to be seen how robust and universal currently fashionable explanations prove to be.

More generally, we started this chapter with a series of questions. Why are some species rare and others common?

in conclusion

Why does a species occur at low population densities in some places and at high densities in others? What factors cause fluctuations in a species’ abundance? Having reached the end of the chapter, it should be clear that none of these questions has a simple answer. We have seen for particular examples why a species is rare, or why another varies in abundance from place to place. But we must not expect the answer to be the same for every species – especially when we start a new study of a species that demands our attention, perhaps by its excessive abundance (a pest) or declining abundance (a target for conservation). It is crucial none the less that we have a clear idea of what the *possible* answers might be and how we might go about obtaining those answers. The aim of this chapter has been to examine those possibilities and how to distinguish them. In the next chapter, we turn explicitly to some of the pressing examples of populations whose abundance we need to understand in order to exert some measure of control – be they pests or natural resources that we wish to exploit.

Summary

We bring together topics from previous chapters, seeking to account for variations in abundance.

Ecologists may emphasize stability or fluctuations. To resolve these contrasting perspectives, it is necessary to distinguish clearly between factors that determine and those that regulate abundance. In doing so, we review historical conflicts between the viewpoints of Nicholson and Andrewartha and Birch. We then outline the demographic, mechanistic and density approaches to the investigation of abundance.

Starting with the demographic approach, we explain key factor analysis, its uses, but also its shortcomings. We therefore also explain λ -contribution analysis, which overcomes some of

the problems with key factor analysis, and in developing this explanation we describe and apply elasticity analysis.

The mechanistic approach relates the level or presence of a factor (amount of food, presence of predators) either to abundance itself or to the population growth rate. This may be simply correlational, but may alternatively involve the experimental perturbation of populations. We note that the introduction of a biological control agent is one particular example of this.

Correlations with density are not absent from other approaches, but the density approach focuses on density dependences in their own right. We explain how time series analyses seek to dissect density dependences, especially the relative strengths of direct and delayed density dependence when abundance at a given point in time is expressed as reflecting abundances at various times in the past ('time lags'). We show, too, how related analyses may be valuable in counting and then characterizing the lags in an optimal description of a time series, and also in evaluating the respective contributions of density-dependent and -independent processes (especially weather) in determining abundance.

Regular, multigeneration cycles have in many ways, and for many years, been the benchmark against which ecologists have tested their ability to understand the determination of abundance. We explain how cycles may be identified within time series and then examine three case studies in detail.

Red grouse cycles illustrate the difficulties of distinguishing between alternative explanations – parasites and kinship/territorial behavior – both of which have support.

Work on cycles in snowshoe hares illustrates the coming together of detailed time series analyses and results obtained by much more direct, experimental means. It also provides a very sobering reminder of the logistical and practical difficulties that need to be accepted and overcome in order to build explanations.

More effort has been expended in studying population cycles in microtine rodents (voles and lemmings) than in any other group of species. We describe geographic trends in cyclicity and the need for an explanation to account for these, and we note that any such explanation must acknowledge that the cycles are the result of a 'second-order' process: a combination of a direct and a delayed density-dependent process. We then examine, in turn, three sets of explanations differing in their source of delayed density dependence: (i) 'intrinsic' theories, including maternal effects; (ii) the 'specialist predation hypothesis', supported by both mathematical models and field experiments, though both of these have also been subject either to criticism or contradictory evidence; and (iii) theories focused on food, which also have their problems.

We conclude by acknowledging that none of the questions posed at the beginning of the chapter has simple answers.

Ecological Applications at the Level of Population Interactions: Pest Control and Harvest Management



15.1 Introduction

Humans are very much a part of all ecosystems. Our activities sometimes motivate us to drive towards extinction the species we identify as pests, to kill individuals of species we harvest for food or fiber while ensuring the persistence of their populations, and to prevent the extinction of species we believe to be endangered. The desired outcomes are very different for pest controllers, harvest managers and conservation ecologists, but all need management strategies based on the theory of population dynamics. Because much of the tool kit developed to manage endangered species is based on the dynamics of individual populations, we dealt with species conservation in Chapter 7 at the end of the first section of the book, which considered the ecology of individual organisms and single species populations. Pest controllers and harvest managers, on the other hand, mostly have to deal explicitly with multispecies interactions, and their work must be informed by the theory concerning population interactions covered in the book's second section (Chapters 8–14). Pest control and harvest management are the topics of the present chapter.

'sustainability' –
an aim of both pest
controllers and
harvest managers

The importance of pest control and harvest management has grown exponentially as the human population has increased (see Section 7.1) and each touches on a different aspect of 'sustainability'. To call an activity 'sustainable' means that it can be continued or repeated for the foreseeable future. Concern has arisen, therefore, precisely because so much human activity is clearly unsustainable. We cannot continue to use the same pesticides if increasing numbers of pests become resistant to them. We cannot (if we wish to have fish to eat in future) continue to remove fish from the sea faster than the remaining fish can replace their lost companions.

Sustainability has thus become one of the core concepts – perhaps the core concept – in an ever-broadening concern for the

fate of the earth and the ecological communities that occupy it. In defining sustainability we used the words 'foreseeable future'. We did so because, when an activity is described as sustainable, it is on the basis of what is known at the time. But many factors remain unknown or unpredictable. Things may take a turn for the worse (as when adverse oceanographic conditions damage a fishery already threatened by overexploitation) or some unforeseen additional problem may be discovered (resistance may appear to some previously potent pesticide). On the other hand, technological advances may allow an activity to be sustained that previously seemed unsustainable (new types of pesticide may be discovered that are more finely targeted on the pest itself rather than innocent bystander species). However, there is a real danger that we observe the many technological and scientific advances that have been made in the past and act on the faith that there will always be a technological 'fix' to solve our present problems, too. Unsustainable practices cannot be accepted simply from faith that future advances will make them sustainable after all.

The recognition of the importance of sustainability as a unifying idea in applied ecology has grown gradually, but there is something to be said for the claim that sustainability really came of age in 1991. This was when the Ecological Society of America published 'The sustainable biosphere initiative: an ecological research agenda', a 'call-to-arms for all ecologists' with a list of 16 co-authors (Lubchenco *et al.*, 1991). And in the same year, the World Conservation Union (IUCN), the United Nations Environment Programme and the World Wide Fund for Nature jointly published *Caring for the Earth. A Strategy for Sustainable Living* (IUCN/UNEP/WWF, 1991). The detailed contents of these documents are less important than their existence. They indicate a growing preoccupation with sustainability, shared by scientists, pressure groups and governments, and recognition that much of what we do is not sustainable. More recently, the emphasis has shifted from a purely ecological perspective to one that incorporates the social and economic conditions influencing sustainability

(Milner-Gulland & Mace, 1998) – this is sometimes referred to as the ‘triple bottomline’ of sustainability.

In this chapter we deal in turn with the application of population theory to the management of pests (Section 15.2) and harvests (Section 15.3). We have seen previously how the details of spatial structuring of populations can affect their dynamics (see Chapters 6 and 14). With this in mind, Section 15.4 presents examples of the application of a metapopulation perspective to pest control and harvest management.

We discussed in Chapter 7 how predicted global climate change is expected to affect species’ distribution patterns. Such conclusions were based on the mapping of species’ fundamental niches onto new global patterns of temperature and rainfall. We will not dwell on this phenomenon in the current chapter, but it should be noted that global change will also impact on population parameters, such as birth and death rates and the timing of breeding (e.g. Walther *et al.*, 2002; Corn, 2003), with implications for the population dynamics of pest and harvested (and endangered) species.

15.2 Management of pests

what is a pest?

A pest species is one that humans consider undesirable. This definition covers a multitude of sins: mosquitoes are pests because they carry diseases or because their bites itch; *Allium* spp. are pests because when harvested with wheat these weeds make bread taste of onions; rats and mice are pests because they feast on stored food; mustellids are pests in New Zealand because they are unwanted invaders that prey upon native birds and insects; garden weeds are pests for esthetic reasons. People want rid of them all.

15.2.1 Economic injury level and economic thresholds

economic injury level defines actual and potential pests

Economics and sustainability are intimately tied together. Market forces ensure that uneconomic practices are not sustainable. One might imagine that the aim of pest control is always total eradication of the pest, but this is not the general rule. Rather, the aim is to reduce the pest population to a level at which it does not pay to achieve yet more control (the economic injury level or EIL). Our discussion here is informed particularly by the theory covered in Chapter 14, which dealt with the combination of factors that determines a species’ average abundance and fluctuations about that average. The EIL for a hypothetical pest is illustrated in Figure 15.1a: it is greater than zero (eradication is not profitable) but it is also below the typical, average abundance of the species. If the species was naturally self-limited to a density below the EIL, then it would never make economic sense to apply ‘control’ measures, and the

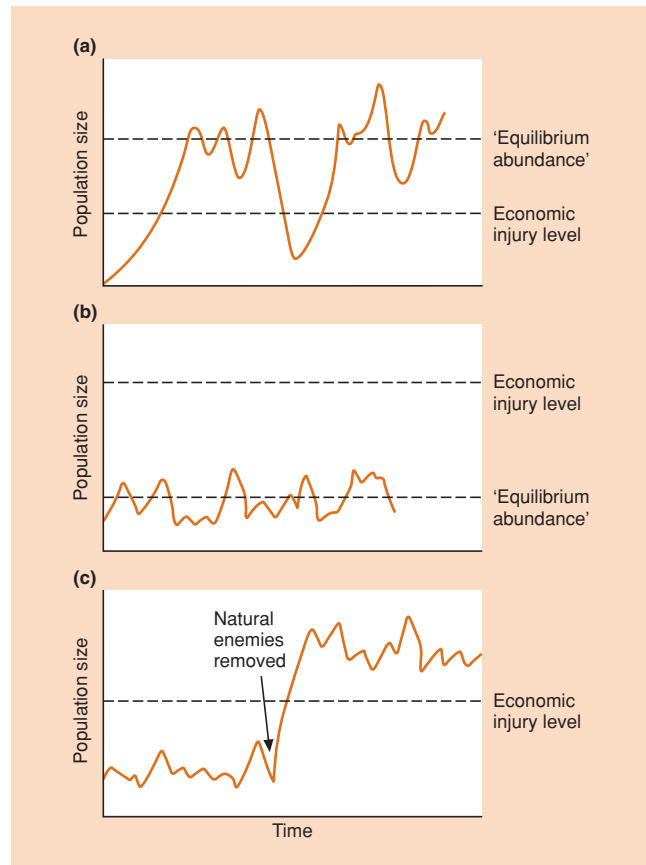


Figure 15.1 (a) The population fluctuations of a hypothetical pest. Abundance fluctuates around an ‘equilibrium abundance’ set by the pest’s interactions with its food, predators, etc. It makes economic sense to control the pest when its abundance exceeds the economic injury level (EIL). Being a pest, its abundance exceeds the EIL most of the time (assuming it is not being controlled). (b) By contrast, a species that cannot be a pest fluctuates always below its EIL. (c) ‘Potential’ pests fluctuate normally below their EIL but rise above it in the absence of one or more of their natural enemies.

species could not, by definition, be considered a ‘pest’ (Figure 15.1b). There are other species, though, that have a carrying capacity in excess of their EIL, but have a typical abundance that is kept below the EIL by natural enemies (Figure 15.1c). These are potential pests. They can become actual pests if their enemies are removed.

When a pest population has reached a density at which it is causing economic injury, however, it is generally too late to start controlling it. More important, then, is the economic threshold (ET):

the economic threshold – getting ahead of the pests

the density of the pest at which action should be taken to prevent it reaching the EIL. ETs are predictions based either on cost–benefit analyses (Ramirez & Saunders, 1999) and detailed studies

of past outbreaks, or sometimes on correlations with climatic records. They may take into account the numbers not only of the pest itself but also of its natural enemies. As an example, in order to control the spotted alfalfa aphid (*Therioaphis trifolii*) on hay alfalfa in California, control measures have to be taken at the times and under the following circumstances (Flint & van den Bosch, 1981):

- 1 In the spring when the aphid population reaches 40 aphids per stem.
- 2 In the summer and fall when the population reaches 20 aphids per stem, but the first three cuttings of hay are not treated if the ratio of ladybirds (beetle predators of the aphids) to aphids is one adult per 5–10 aphids or three larvae per 40 aphids on standing hay or one larva per 50 aphids on stubble.
- 3 During the winter when there are 50–70 aphids per stem.

15.2.2 Chemical pesticides, target pest resurgence and secondary pests

Chemical pesticides are a key part of the armory of pest managers but they have to be used with care because population theory (see, in particular, Chapter 14) predicts some undesirable responses to the application of a pesticide. Below we discuss the range of chemical pesticides and herbicides before proceeding to consider some undesirable consequences of their use.

15.2.2.1 Insecticides

insecticides and how they work

The use of *inorganics* goes back to the dawn of pest control and, along with the botanicals (below), they were the chemical weapons of the expanding army of insect pest managers of the 19th and early 20th century. They are usually metallic compounds or salts of copper, sulfur, arsenic or lead – and are primarily stomach poisons (i.e. they are ineffective as contact poisons) and they are therefore effective only against insects with chewing mouthparts. This, coupled with their legacy of persistent, broadly toxic metallic residues, has led now to their virtual abandonment (Horn, 1988).

Naturally occurring insecticidal plant products, or *botanicals*, such as nicotine from tobacco and pyrethrum from chrysanthemums, having run a course similar to the inorganics, have now also been largely superseded, particularly because of their instability on exposure to light and air. However, a range of *synthetic pyrethroids*, with much greater stability, such as permethrin and deltamethrin, have replaced other types of organic insecticide (described below) because of their relative selectivity against pests as opposed to beneficial species (Pickett, 1988).

Chlorinated hydrocarbons are contact poisons that affect nerve-impulse transmission. They are insoluble in water but show a high

affinity for fats, thus tending to become concentrated in animal fatty tissue. The most notorious is DDT: a Nobel Prize was awarded for its rediscovery in 1948, but it was suspended from all but emergency uses in the USA in 1973 (although it is still being used in poorer countries). Others in use are toxaphene, aldrin, dieldrin, lindane, methoxychlor and chlordane.

Organophosphates are also nerve poisons. They are much more toxic (to both insects and mammals) than the chlorinated hydrocarbons, but are generally less persistent in the environment. Examples are malathion, parathion and diazinon.

Carbamates have a mode of action similar to the organophosphates, but some have a much lower mammalian toxicity. However, most are extremely toxic to bees (necessary for pollination) and parasitic wasps (the likely natural enemies of insect pests). The best-known carbamate is carbaryl.

Insect growth regulators are chemicals of various sorts that mimic natural insect hormones and enzymes, and hence interfere with normal insect growth and development. As such, they are generally harmless to vertebrates and plants, although they may be as effective against a pest's natural insect enemies as against the pest itself. The two main types that have been used effectively to date are: (i) chitin-synthesis inhibitors such as diflubenzuron, which prevent the formation of a proper exoskeleton when the insect molts; and (ii) juvenile hormone analogs such as methoprene, which prevent pest insects from molting into their adult stage, and hence reduce the population size in the next generation.

Semiochemicals are not toxins but chemicals that elicit a change in the behavior of the pest (literally 'chemical signs'). They are all based on naturally occurring substances, although in a number of cases it has been possible to synthesize either the semiochemicals themselves or analogs of them. Pheromones act on members of the same species; allelochemicals on members of another species. Sex-attractant pheromones are used commercially to control pest moth populations by interfering with mating (Reece, 1985), whilst the aphid alarm pheromone is used to enhance the effectiveness of a fungal pathogen against pest aphids in glasshouses in Great Britain by increasing the mobility of the aphids, and hence their rate of contact with fungal spores (Hockland *et al.*, 1986). These semiochemicals, along with the insect growth regulators, are sometimes referred to as 'third-generation' insecticides (following the inorganics and the organic toxins). Their development is relatively recent (Forrester, 1993).

15.2.2.2 Herbicides

Here, too, *inorganics* were once important although they have mostly been replaced, largely owing to the combined problems of persistence and nonspecificity. However, for these very reasons, borates for example, absorbed by plant roots and translocated to above-ground parts, are still sometimes used to provide semipermanent sterility to areas where no vegetation

the tool-kit of herbicides

of any sort is wanted. Others include a range of arsenicals, ammonium sulfamate and sodium chlorate (Ware, 1983).

More widely used are the *organic arsenicals*, for instance disodium methylarsonate. These are usually applied as spot treatments (since they are nonselective) after which they are translocated to underground tubers and rhizomes where they disrupt growth.

By contrast, the highly successful *phenoxy* or *hormone* weed killers, translocated throughout the plant, tend to be very much more selective. For instance, 2,4-D is highly selective against broad-leaved weeds, whilst 2,4,5-trichlorophenoxyethanoic acid (2,4,5-T) is used mainly to control woody perennials. They appear to act by inhibiting the production of enzymes needed for coordinated plant growth, leading ultimately to plant death.

The *substituted amides* have diverse biological properties. For example, diphenamid is largely effective against seedlings rather than established plants, and is therefore applied to the soil around established plants as a 'pre-emergence' herbicide, preventing the subsequent appearance of weeds. Propanil, on the other hand, has been used extensively on rice fields as a selective post-emergence agent.

The *nitroanilines* (e.g. trifluralin) are another group of soil-incorporated pre-emergence herbicides in very widespread use. They act, selectively, by inhibiting the growth of both roots and shoots.

The *substituted ureas* (e.g. monuron) are mostly rather nonselective pre-emergence herbicides, although some have post-emergence uses. Their mode of action is to block electron transport.

The *carbamates* were described amongst the insecticides, but some are herbicides, killing plants by stopping cell division and plant tissue growth. They are primarily selective, pre-emergence weed killers. One example, asulam, is used mostly for grass control amongst crops, and is also effective in reforestation and Christmas tree plantings.

The *thiocarbamates* (e.g. S-ethyl dipropylthiocarbamate) are another group of soil-incorporated pre-emergence herbicides, selectively inhibiting the growth of roots and shoots that emerge from weed seeds.

Amongst the *heterocyclic nitrogen* herbicides, probably the most important are the *triazines* (e.g. metribuzin). These are effective blockers of electron transport, mostly used for their post-emergence activity.

The *phenol derivatives*, particularly the nitrophenols such as 2-methyl-4,6-dinitrophenol, are contact chemicals with broad-spectrum toxicity extending beyond plants to fungi, insects and mammals. They act by uncoupling oxidative phosphorylation.

The *bipyridyliums* contain two important herbicides, diquat and paraquat. These are powerful, very fast acting contact chemicals of widespread toxicity that act by the destruction of cell membranes.

Finally worth mentioning is glyphosate (a *glyphosphate* herbicide): a nonselective, nonresidual, translocated, foliar-applied chemical, popular for its activity at any stage of plant growth and at any time of the year.

15.2.2.3 Target pest resurgence

A pesticide gets a bad name if, as is usually the case, it kills more species than just the one at which it is aimed. However, in the context of the sustainability of agriculture, the bad name is especially justified if it kills the pests' natural enemies and so contributes to undoing what it was employed to do. Thus, the numbers of a pest sometimes increase rapidly some time after the application of a pesticide. This is known as 'target pest resurgence' and occurs when the treatment kills both large numbers of the pest *and* large numbers of its natural enemies (an example is presented below in Figure 15.2). Pest individuals that survive the pesticide or that migrate into the area later find themselves with a plentiful food resource but few, if any, natural enemies. The pest population may then explode. Populations of natural enemies will probably eventually re-establish but the timing depends both on the relative toxicity of the pesticide to target and nontarget species and the persistence of the pesticide in the environment, something that varies dramatically from one pesticide to another (Table 15.1).

the pest bounces back because its enemies are killed

	Toxicity				Persistence
	Rat	Fish	Bird	Honeybee	
Permethrin (pyrethroid)	2	4	2	5	2
DDT (organochlorine)	3	4	2	2	5
Lindane (organochlorine)	3	3	2	4	4
Ethyl parathion (organophosphate)	5	2	5	5	2
Malathion (organophosphate)	2	2	1	4	1
Carbaryl (carbamate)	2	1	1	4	1
Diflubenzuron (chitin-synthesis inhibitor)	1	1	1	1	4
Methoprene (juvenile hormone analogue)	1	1	1	2	2
<i>Bacillus thuringiensis</i>	1	1	1	1	1

Table 15.1 The toxicity to nontarget organisms, and the persistence, of selected insecticides. Possible ratings range from a minimum of 1 (which may, therefore, include zero toxicity) to a maximum of 5. Most damage is done by insecticides that combine persistence with acute toxicity to nontarget organisms. This clearly applies, to an extent, to each of the first six (broad-spectrum) insecticides. (After Metcalf, 1982; Horn, 1988.)

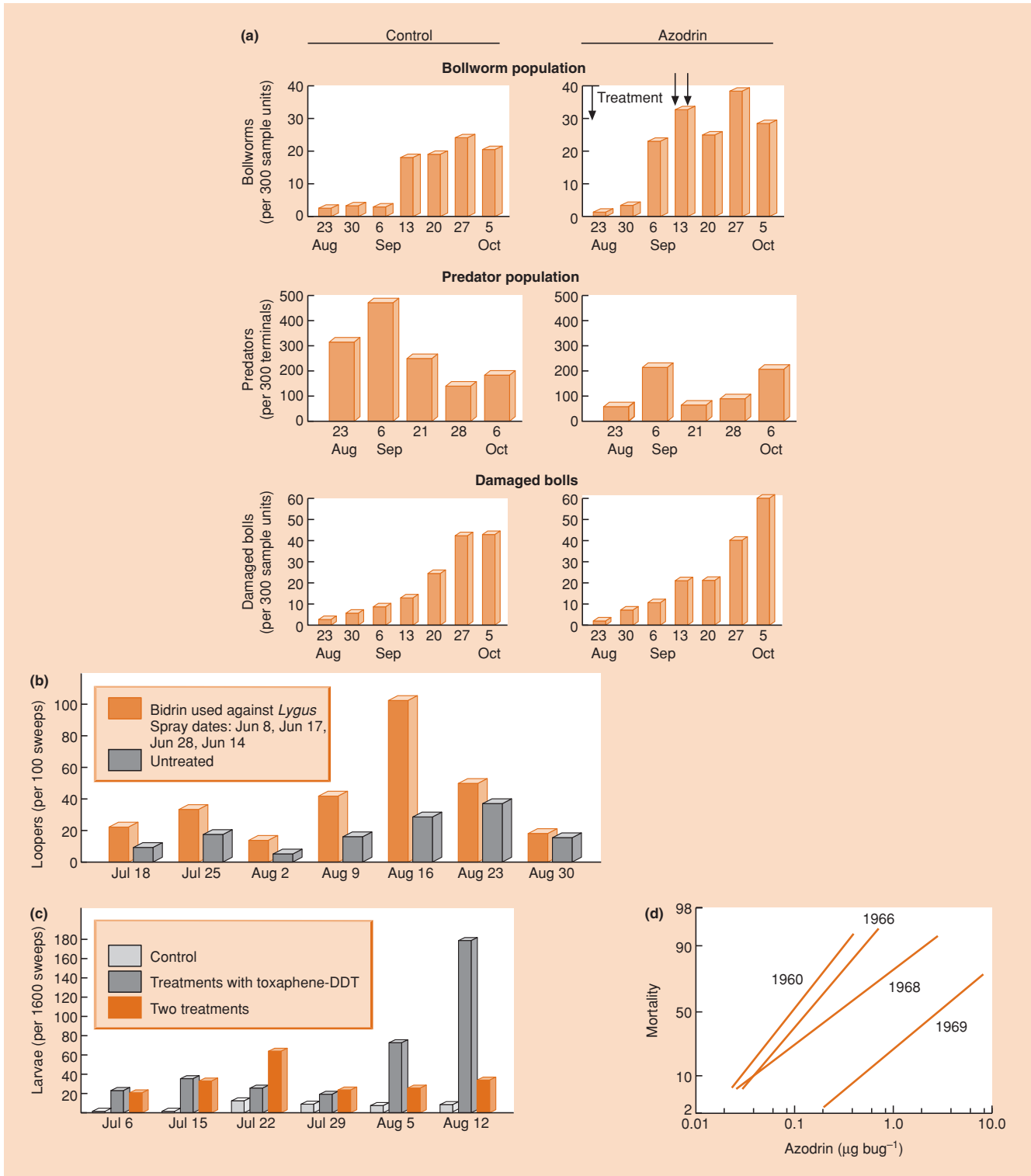


Figure 15.2 Pesticide problems amongst cotton pests in the San Joaquin Valley, California. (a) Target pest resurgence: cotton bollworms (*Heliothis zea*) resurged because the abundance of their natural predators was reduced – the number of damaged bolls was higher. (b) An increase in cabbage loopers (*Trichoplusia ni*) and (c) in beet army worms (*Spodoptera exigua*) were seen when plots were sprayed against the target lygus bugs (*Lygus hesperus*) – both are examples of secondary pest outbreaks. (d) Increasing resistance of lygus bugs to Azodrin®. (After van den Bosch *et al.*, 1971.)

15.2.2.4 Secondary pests

nonpests become pests when their enemies and competitors are killed

The after-effects of a pesticide may involve even more subtle reactions. When a pesticide is applied, it may not be only the target pest that resurges. Alongside the target are likely to be a number of potential pest species that had been kept in check by their natural enemies (see Figure 15.1c). If the pesticide destroys these, the potential pests become real ones – and are called secondary pests. A dramatic example concerns the insect pests of cotton in the southern part of the USA. In 1950, when mass dissemination of organic insecticides began, there were two primary pests: the Alabama leafworm and the boll weevil (*Anthonomus grandis*), an invader from Mexico (Smith, 1998). Organochlorine and organophosphate insecticides (see Section 15.2.2.1) were applied fewer than five times a year and initially had apparently miraculous results – cotton yields soared. By 1955, however, three secondary pests had emerged: the cotton bollworm, the cotton aphid and the false pink bollworm. The pesticide applications rose to 8–10 per year. This reduced the problem of the aphid and the false pink bollworm, but led to the emergence of five further secondary pests. By the 1960s, the original two pest species had become eight and there were, on average, an unsustainable 28 applications of insecticide per year. A study in the San Joaquin Valley, California, revealed target pest resurgence (in this case cotton bollworm was the target species; Figure 15.2a) and secondary pest outbreaks in action (cabbage loopers and beet army worms increased after insecticide application against another target species, the lygus bug; Figure 15.2b, c). Improved performance in pest management will depend on a thorough understanding of the interactions amongst pests and nonpests as well as detailed knowledge, through testing, of the action of potential pesticides against the various species.

mortality of nontarget species in general

Sometimes the unintended effects of pesticide application have been much less subtle than target pest or secondary pest resurgence. The potential for disaster is illustrated by the occasion when massive doses of the insecticide dieldrin were applied to large areas of Illinois farmland from 1954 to 1958 to ‘eradicate’ a grassland pest, the Japanese beetle. Cattle and sheep on the farms were poisoned, 90% of cats and a number of dogs were killed, and among the wildlife 12 species of mammals and 19 species of birds suffered losses (Luckman & Decker, 1960). Outcomes such as this argue for a precautionary approach in any pest management exercise. Coupled with much improved knowledge about the toxicity and persistence of pesticides, and the development of more specific and less persistent pesticides, such disasters should never occur again.

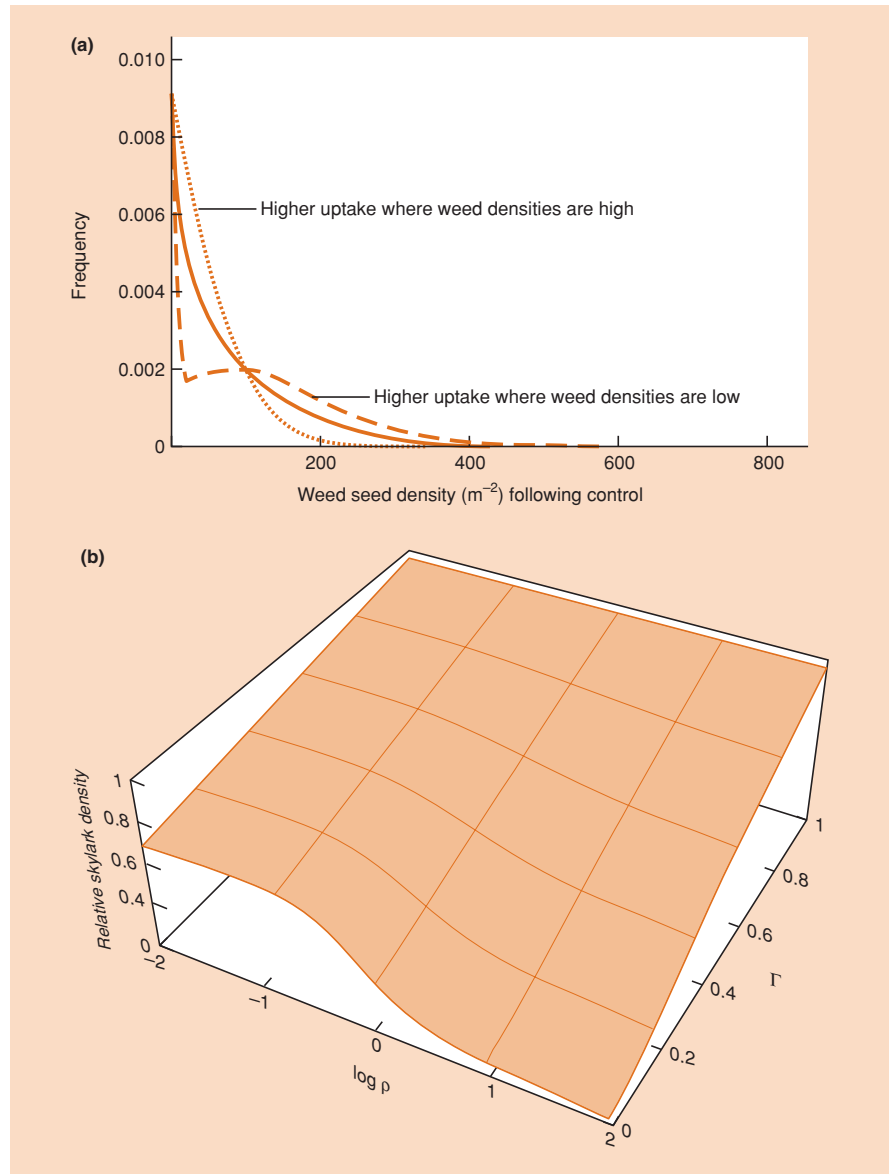
15.2.3 Herbicides, weeds and farmland birds

Herbicides are used in very large amounts and on a worldwide scale. They are active against pest plants and when used at commercial rates appear to have few significant effects on animals. Herbicide pollution of the environment did not, until relatively recently, arouse the passions associated with insecticides. However, conservationists now worry about the loss of ‘weeds’ that are the food hosts for larvae of butterflies and other insects and whose seeds form the main diet of many birds. A recent development has been the genetic modification of crops such as sugar beet to produce resistance to the nonselective herbicide glyphosate (see Section 15.2.2.2). This allows the herbicide to be used to effectively control weeds that normally compete with the crop without adverse affect on the sugar beet itself.

unintended effects of the genetic modification of crops with herbicide resistance

Fat hen (*Chenopodium album*), a plant that occurs worldwide, is one weed that can be expected to be affected adversely by the farming of genetically modified (GM) crops; but the seeds of fat hen are an important winter food source for farmland birds, including the skylark (*Alauda arvensis*). Watkinson *et al.* (2000) took advantage of the fact that the population ecologies of both fat hen and skylarks have been intensively studied and incorporated both into a model of the impacts of GM sugar beet on farmland populations. Skylarks forage preferentially in weedy fields and aggregate locally in response to weed seed abundance. Hence, the impact of GM sugar beet on the birds will depend critically on the extent to which high-density patches of weeds are affected. Watkinson *et al.* incorporated the possible effects of weed seed density on farming practice. Their model assumed: (i) that before the introduction of GM technology, most farms have a relatively low density of weed seeds, with a few farms having very high densities (solid line in Figure 15.3a); and (ii) the probability of a farmer adopting GM crops is related to seed bank density through a parameter ρ . Positive values of ρ mean that farmers are more likely to adopt the technology where seed densities are currently high and there is the potential to reduce yield losses to weeds. This leads to an increase in the relative abundance of low-density fields (dotted line in Figure 15.3a). Negative values of ρ indicate that farmers are more likely to adopt the technology where seed densities are currently low (intensively managed farms), perhaps because a history of effective weed control is correlated with a willingness to adopt new technology. This leads to a decreased frequency of low-density fields (dashed line in Figure 15.3a). Note that ρ is not an ecological parameter. Rather it reflects a socioeconomic response to the introduction of new technology. The way that farmers will respond is not self-evident and needs to be included as a variable in the model. It turns out that the relationship between current weed levels and uptake of the new technology (ρ) is as important to bird population

Figure 15.3 (a) Frequency distributions of mean seed densities across farms before the introduction of GM sugar beet (solid line), and in two situations where the technology has been adopted: where the technology is preferentially adopted on farms where weed density is currently high (dotted line) and where it is currently low (dashed line). (b) The relative density of skylarks in fields in winter (vertical axis; unity indicates field use before the introduction of GM crops) in relation to ρ (horizontal axis; positive values mean farmers are more likely to adopt GM technology where seed densities are currently high, negative values where seed densities are currently low) and to the approximate reduction in weed seed bank density due to the introduction of GM crops (Γ , third axis; realistic values are those less than 0.1). Note that the parameter space that real systems are expected to occupy is the ‘slice’ of the diagram nearest to you, where small positive or negative values of ρ give quite different skylark densities. (After Watkinson *et al.*, 2000.)



density as the direct impact of the technology on weed abundance (Figure 15.3b), emphasizing the need for resource managers to think in terms of the triple bottomline of sustainability, with its ecological, social and economic dimensions.

15.2.4 Evolution of resistance to pesticides

evolved resistance: a widespread problem

Chemical pesticides lose their role in sustainable agriculture if the pests evolve resistance. The evolution of pesticide resistance is simply natural selection in action. It is almost cer-

tain to occur when vast numbers of individuals in a genetically variable population are killed in a systematic way by the pesticide. One or a few individuals may be unusually resistant (perhaps because they possess an enzyme that can detoxify the pesticide). If the pesticide is applied repeatedly, each successive generation of the pest will contain a larger proportion of resistant individuals. Pests typically have a high intrinsic rate of reproduction, and so a few individuals in one generation may give rise to hundreds or thousands in the next, and resistance spreads very rapidly in a population.

This problem was often ignored in the past, even though the first case of DDT resistance was reported as early as 1946

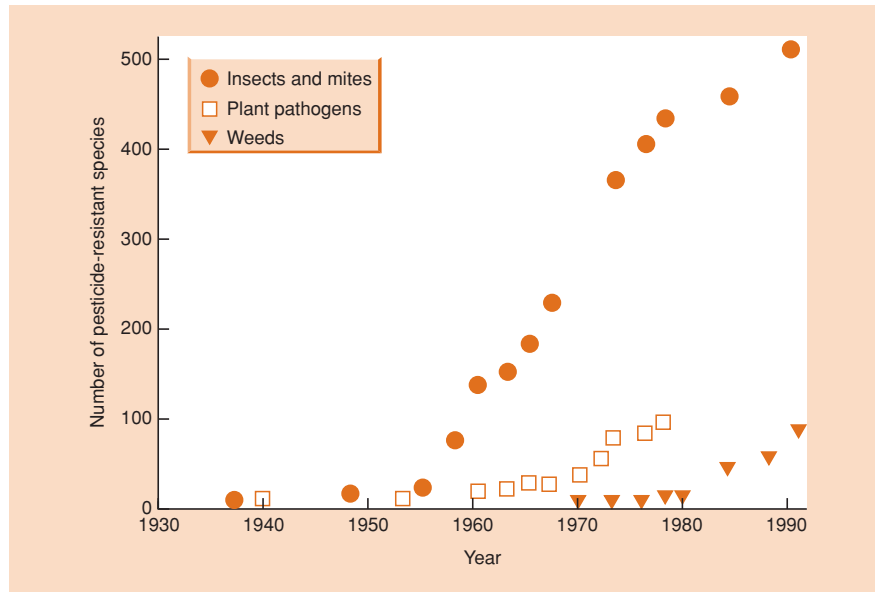


Figure 15.4 The increase in the number of arthropod (insects and mites), plant pathogens and weed species reported to be resistant to at least one pesticide. (After Gould, 1991.)

(in house-flies, *Musca domestica*, in Sweden). The scale of the problem is illustrated in Figure 15.4, which shows the exponential increases in the number of invertebrates, weeds and plant pathogens resistant to insecticides. The cotton pest study described earlier also provides evidence of the evolution of resistance to a pesticide (see Figure 15.2d). Even rodents and rabbits (*Oryctolagus cuniculus*) have evolved resistance to certain pesticides (Twigg *et al.*, 2002).

managing resistance

The evolution of pesticide resistance can be slowed, though, by changing from one pesticide to another, in a repeated sequence that is rapid enough that resistance does not have time to emerge (Roush & McKenzie, 1987). River blindness, a devastating disease that has now been effectively eradicated over large areas of Africa, is transmitted by the biting blackfly *Simulium damnosum*, whose larvae live in rivers. A massive helicopter pesticide spraying effort in several African countries (50,000 km of river were being treated weekly by 1999; Yameogo *et al.*, 2001) began with Temephos, but resistance appeared within 5 years (Table 15.2). Temephos was then replaced by another organophosphate, Chlorphoxim, but resistance rapidly evolved to this too. The strategy of using a range of pesticides on a rotational basis has prevented further evolution of resistance and by 1994 there were few populations that were still resistant to Temephos (Davies, 1994).

If chemical pesticides brought nothing but problems, however – if their use was intrinsically and acutely unsustainable – then they would already have fallen out of widespread use. This has not happened. Instead, their rate of production has increased rapidly. The ratio of cost to benefit for the individual producer has generally remained in favor of pesticide use. Moreover, in many poorer countries, the prospect of imminent mass starvation, or

Table 15.2 History of pesticide use against the aquatic larvae of blackflies, the vectors of river blindness in Africa. After early concentration on Temephos and Chlorphoxim, to which the insects became resistant, pesticides were used on a rotational basis to prevent the evolution of resistance. (After Davies, 1994.)

Name of pesticide	Class of chemical	History of use
Temephos	Organophosphate	1975 to present
Chlorphoxim	Organophosphate	1980–90
<i>Bacillus thuringiensis</i> H14	Biological insecticide	1980 to present
Permethrin	Pyrethroid	1985 to present
Carbosulfan	Carbamate	1985 to present
Pyraclufos	Organic phosphate	1991 to present
Phoxim	Organophosphate	1991 to present
Etofenprox	Pyrethroid	1994 to present

of an epidemic disease, are so frightening that the social and health costs of using pesticides have to be ignored. In general the use of pesticides is justified by objective measures such as ‘lives saved’, ‘economic efficiency of food production’ and ‘total food produced’. In these very fundamental senses, their use may be described as sustainable. In practice, sustainability depends on continually developing new pesticides that keep at least one step ahead of the pests: pesticides that are less persistent, biodegradable and more accurately targeted at the pests.

15.2.5 Biological control

Outbreaks of pests occur repeatedly and so does the need to apply pesticides. But biologists can sometimes replace chemicals by

biological control:
the use of natural
enemies in a variety
of ways

another tool that does the same job and often costs a great deal less – biological control (the manipulation of the natural enemies of pests). Biological control involves the application of theory about interactions between species

and their natural enemies (see Chapters 10, 12 and 14) to limit the population density of specific pest species. There are a variety of categories of biological control.

The first is the *introduction* of a natural enemy from another geographic area – very often the area in which the pest originated prior to achieving pest status – in order that the control agent should persist and thus maintain the pest, long term, below its economic threshold. This is a case of a desired invasion of an exotic species and is often called *classical biological control* or *importation*.

By contrast, *conservation biological control* involves manipulations that augment the density or persistence of populations of generalist natural enemies that are native to the pest’s new area (Barbosa, 1998).

Inoculation is similar to introduction, but requires the periodic release of a control agent where it is unable to persist throughout the year, with the aim of providing control for only one or perhaps a few generations. A variation on the theme of inoculation is ‘augmentation’, which involves the release of an indigenous natural enemy in order to supplement an existing population, and is also therefore carried out repeatedly, typically to intercept a period of rapid pest population growth.

Finally, *inundation* is the release of large numbers of a natural enemy, with the aim of killing those pests present at the time, but with no expectation of providing long-term control as a result of the control agent’s population increasing or maintaining itself. By analogy with the use of chemicals, agents used in this way are referred to as biological pesticides.

Insects have been the main agents of biological control against both insect pests (where parasitoids have been particularly useful) and weeds. Table 15.3 summarizes the extent to which they have been used and the proportion of cases where the establishment of an agent has greatly reduced or eliminated the need for other control measures (Waage & Greathead, 1988).

Table 15.3 The record of insects as biological control agents against insect pests and weeds. (After Waage & Greathead, 1988.)

	<i>Insect pests</i>	<i>Weeds</i>
Control agent species	563	126
Pest species	292	70
Countries	168	55
Cases where agent has become established	1063	367
Substantial successes	421	113
Successes as a percentage of establishments	40	31

Probably the best example of ‘classical’ biological control is itself a classic. Its success marked the start of biological control in a modern sense.

The cottony cushion scale insect, *Icerya purchasi*, was first discovered as a pest of Californian citrus orchards in 1868. By 1886 it had brought the citrus industry close to the point of destruction. Ecologists initiated a worldwide correspondence to try and discover the natural home and natural enemies of the scale, eventually leading to the importation to California of about 12,000 *Cryptochaetum* (a dipteran parasitoid) from Australia and 500 predatory ladybird beetles (*Rodolia cardinalis*) from Australia and New Zealand. Initially, the parasitoids seemed simply to have disappeared, but the predatory beetles underwent such a population explosion that all infestations of the scale insects in California were controlled by the end of 1890. Although the beetles have usually taken most or all of the credit, the long-term outcome has been that the beetles are instrumental in keeping the scale in check inland, but *Cryptochaetum* is the main agent of control on the coast (Flint & van den Bosch, 1981).

This example illustrates a number of important general points. Species may become pests simply because, by colonization of a new area, they escape the control of their natural enemies (the enemy release hypothesis) (Keane & Crawley, 2002). Biological control by importation is thus, in an important sense, restoration of the status quo for the specific predator–prey interaction (although the overall ecological context is certain to differ from what would have been the case where the pest and control agent originated). Biological control requires the classical skills of the taxonomist to find the pest in its native habitat, and particularly to identify and isolate its natural enemies. This may often be a difficult task – especially if the natural enemy has the desired effect of keeping the target species at a low carrying capacity, since both the target and the agent will then be rare in their natural habitat. Nevertheless, the rate of return on investment can be highly favorable. In the case of the cottony cushion scale, biological control has subsequently been transferred to 50 other countries and savings have been immense. In addition, this example illustrates the importance of establishing several, hopefully complementary, enemies to control a pest. Finally, classical biological control, like natural control, can be destabilized by chemicals. The first use of DDT in Californian citrus orchards in 1946–47 against the citricola scale *Coccus pseudomagnoliarum* led to an outbreak of the (by then) rarely seen cottony cushion scale when the DDT almost eliminated the ladybirds. The use of DDT was terminated.

Many pests have a diversity of natural enemies that already occur in their vicinity. For example, the aphid pests of wheat (e.g. *Sitobion avenae* or *Rhopalosiphum* spp.) are attacked by

cottony cushion scale
insect: a classic case
of importation . . .

. . . illustrating
several general points

conservation
biological control
of wheat aphids

coccinellid and other beetles, heteropteran bugs, lacewings (Chrysopidae), syrphid fly larvae and spiders – all part of a large group of specialist aphid predators and generalists that include them in their diet (Brewer & Elliott, 2004). Many of these natural enemies overwinter in the grassy boundaries at the edge of wheat fields, from where they disperse and reduce aphid populations around the field edges. The planting of grassy strips within the fields can enhance these natural populations and the scale of their impact on aphid pests. This is an example of ‘conservation biological control’ in action (Barbosa, 1998).

inoculation against
glasshouse pests

‘Inoculation’ as a means of biological control is widely practised in the control of arthropod pests in glasshouses, a situation in which crops are removed, along with the pests and their natural enemies, at the end of the growing season (van Lenteren & Woets, 1988). Two particularly important species of natural enemy used in this way are *Phytoseiulus persimilis*, a mite that preys on the spider mite *Tetranychus urticae*, a pest of cucumbers and other vegetables, and *Encarsia formosa*, a chalcid parasitoid wasp of the whitefly *Trialeurodes vaporariorum*, a pest in particular of tomatoes and cucumbers. By 1985 in Western Europe, around 500 million individuals of each species were being produced each year.

microbial control
of insects via
inundation

‘Inundation’ often involves the use of insect pathogens to control insect pests (Payne, 1988). By far the most widespread and important agent is the bacterium *Bacillus thuringiensis*, which can easily be produced on artificial media. After being ingested by insect larvae, gut juices release powerful toxins and death occurs 30 min to 3 days later. Significantly, there is a range of varieties (or ‘pathotypes’) of *B. thuringiensis*, including one specific against lepidoptera (many agricultural pests), another against diptera, especially mosquitos and blackflies (the vectors of malaria and onchocerciasis) and a third against beetles (many agricultural and stored product pests). *B. thuringiensis* is used inundatively as a microbial insecticide. Its advantages are its powerful toxicity against target insects and its lack of toxicity against organisms outside this narrow group (including ourselves and most of the pest’s natural enemies). Plants, including cotton (*Gossypium hirsutum*), have been genetically modified to express the *B. thuringiensis* toxin (insecticidal crystal protein Cry1Ac). The survivorship of pink bollworm larvae (*Pectinaphora gossypiella*) on genetically modified cotton was 46–100% lower than on nonmodified cotton (Lui *et al.*, 2001). Concern has arisen about the widespread insertion of Bt into commercial genetically modified crops, because of the increased likelihood of the development of resistance to one of the most effective ‘natural’ insecticides available.

biological control
is not always
environmentally
friendly

Biological control may appear to be a particularly environmentally friendly

approach to pest control, but examples are coming to light where even carefully chosen and apparently successful introductions of biological control agents have impacted on nontarget species. For example, a seed-feeding weevil (*Rhinocyllus conicus*), introduced to North America to control exotic *Carduus* thistles, attacks more than 30% of native thistles (of which there are more than 90 species), reducing thistle densities (by 90% in the case of the Platte thistle *Cirsium canescens*) with consequent adverse impacts on the populations of a native picture-winged fly (*Paracantha culta*) that feeds on thistle seeds (Louda *et al.*, 2003a). Louda *et al.* (2003b) reviewed 10 biological control projects that included the unusual but worthwhile step of monitoring nontarget effects and concluded that relatives of the target species were most likely to be attacked whilst rare native species were particularly susceptible. Their recommendations for management included the avoidance of generalist control agents, an expansion of host-specificity testing and the need to incorporate more ecological information when evaluating potential biological control agents.

15.2.6 Integrated pest management

A variety of management implications of our understanding of pest population dynamics have been presented in previous sections. However, it is important to take a broader perspective and consider how all the different tools at the pest controller’s disposal

IPM: an ecologically
rather than
chemically based
philosophy

can be deployed most effectively, both to maximize the economic benefit of reducing pest density and to minimize the adverse environmental and health consequences. This is what integrated pest management (IPM) is intended to achieve. It combines physical control (for example, simply keeping invaders from arriving, keeping pests away from crops, or picking them off by hand when they arrive), cultural control (for example, rotating the crops planted in a field so pests cannot build up their numbers over several years), biological and chemical control, and the use of resistant varieties of crop. IPM came of age as part of the reaction against the unthinking use of chemical pesticides in the 1940s and 1950s.

IPM is ecologically based and relies heavily on natural mortality factors, such as weather and enemies, and seeks to disrupt the latter as little as possible. It aims to control pests below the EIL, and it depends on monitoring the abundance of pests and their natural enemies and using various control methods as complementary parts of an overall program. Broad-spectrum pesticides in particular, although not excluded, are used only very sparingly, and if chemicals are used at all it is in ways that minimize the costs and quantities used. The essence of the IPM approach is to make the control measures fit the pest problem, and no two problems are the same – even in adjacent fields. Thus, IPM often involves the development of computer-based expert systems

that can be used by farmers to diagnose pest problems and suggest appropriate responses (Mahaman *et al.*, 2003).

IPM for the potato tuber moth

The caterpillar of the potato tuber moth (*Phthorimaea operculella*) commonly damages crops in New Zealand. An invader from a warm temperate

subtropical country, it is most devastating when conditions are warm and dry (i.e. when the environment coincides closely with its optimal niche requirements – see Chapter 3). There can be as many as 6–8 generations per year and different generations mine leaves, stems and tubers. The caterpillars are protected both from natural enemies (parasitoids) and insecticides when in the tuber, so control must be applied to the leaf-mining generations. The IPM strategy for potato tuber moth (Herman, 2000) involves: (i) monitoring (female pheromone traps, set weekly from mid summer, are used to attract males, which are counted); (ii) cultural methods (the soil is cultivated to prevent soil cracking, soil ridges are molded up more than once and soil moisture is maintained); and (iii) the use of insecticides, but only when absolutely necessary (most commonly the organophosphate, methamidophos). Farmers follow the decision tree shown in Figure 15.5.

integration of IPM in sustainable farming systems

Implicit in the philosophy of IPM is the idea that pest control cannot be isolated from other aspects of food production and it is especially bound up with the means by which soil fertility

is maintained and improved. These broader sustainable agricultural systems, including IFS (integrated farming systems) in the USA and LIFE (lower input farming and environment) in Europe (International Organisation for Biological Control, 1989; National Research Council, 1990), have advantages in terms of reduced environmental hazards. Even so, it is unreasonable to suppose that they will be adopted widely unless they are also sound in economic terms. In this context, Figure 15.6 shows the yields of apples from organic, conventional and integrated production systems in Washington State from 1994 to 1999 (Reganold *et al.*, 2001).

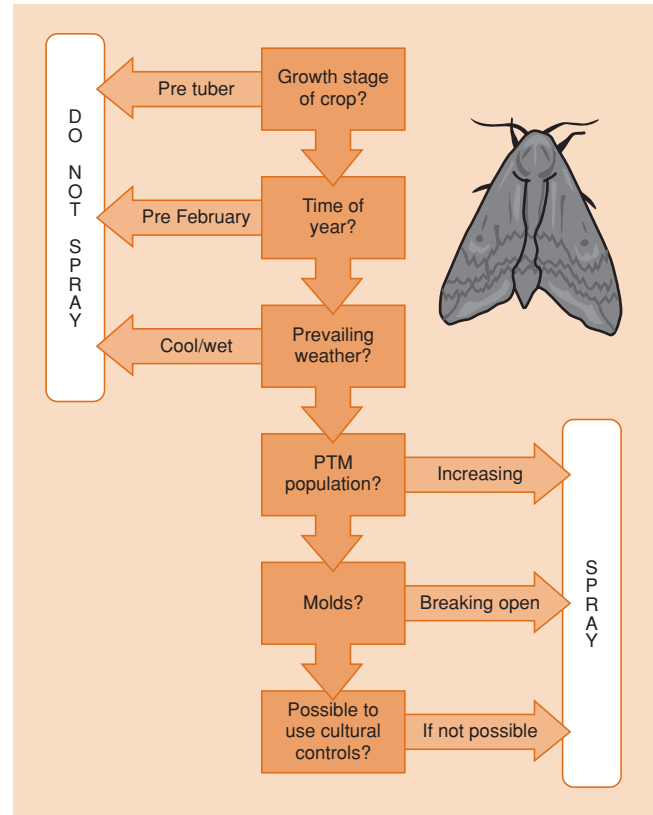


Figure 15.5 Decision flow chart for the integrated pest management of potato tuber moths (PTM) in New Zealand. Boxed phrases are questions (e.g. ‘what is the growth stage of the crop?’), the words in the arrows are the farmer’s answers to the questions (e.g. ‘before the tuber has formed’) and the recommended action is shown in the vertical box (‘don’t spray the crop’). Note that February is late summer in New Zealand. (After Herman, 2000.) Photograph © International Potato Center (CIP).

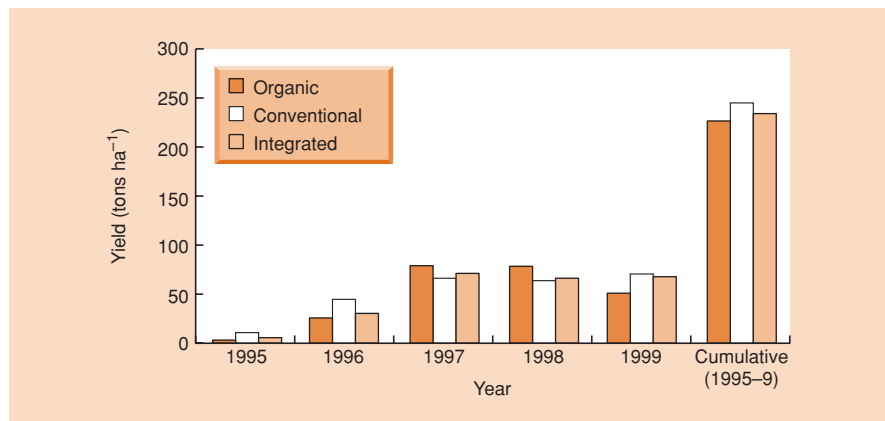


Figure 15.6 The fruit yields of three apple production systems. (From Reganold *et al.*, 2001.)

Organic management excludes such conventional inputs as synthetic pesticides and fertilizers whilst integrated farming uses reduced amounts of chemicals by integrating organic and conventional approaches. All three systems gave similar apple yields but the organic and integrated systems had higher soil quality and potentially lower environmental impacts. When compared with conventional and integrated systems, the organic system produced sweeter apples, higher profitability and greater energy efficiency. Note, however, that despite some widely held beliefs, organic farming is not totally free of adverse environmental consequences. For example, some approved pesticides are just as harmful as synthetic ones whilst the application of animal manure may lead to undesirable levels of nitrate runoff to streams just as synthetic fertilizers can (Trewavas, 2001). There is a need for research to compare the types and magnitudes of environmental consequences of the various approaches to agricultural management.

15.2.7 The importance of the early control of invaders

when a new pest invades . . .

Many pests begin life as exotic invaders. The best way to deal with the problem of potential invaders is to understand their immigration potential (see Section 7.4.2) and prevent their arrival by careful biosecurity processes at a nation's point of entry, or elsewhere on trade routes (Wittenberg & Cock, 2001). However, there are so many potential invaders that it is unrealistic to expect that they all will be prevented from arriving. Moreover, many arrivals will not establish, and many of those that do establish will do so without dramatic ecological consequences. Managers need to focus on the really problematic cases. Thus, the next step in an invader management strategy is to prioritize those that might arrive (or that have recently arrived) according to their likelihood of persisting, establishing large populations, spreading through the new area and causing significant problems. This is not an easy matter, but particular life history traits provide useful pointers (dealt with in Section 7.3.2). We will see in Chapter 22 that assessment of the potential to do harm at higher ecological levels (community/ecosystem) can also be helpful in prioritizing invaders for special attention (see Section 22.3.1).

. . . early control is best

The arrival of an exotic species with a high likelihood of becoming a significant invasive species should be a matter for urgent action, because this is the stage at which eradication is both feasible and easy to justify economically. Such campaigns sometimes rely on fundamental knowledge of population ecology. An example is the eradication of the South African sabellid polychaete worm, *Terebrasabella heterouncinata*, a parasite of abalone and other gastropods that became established near the outflow of an abalone aquaculture facility in California (Culver & Kuris, 2000).

Its population biology was understood sufficiently to know it was specific to gastropods, that two species of *Tegula* were its principal hosts in the area, and that large snails were most susceptible to the parasite. Volunteers removed 1.6 million large hosts, thereby reducing the density of susceptible hosts below that needed for parasite transmission (see Chapter 12), which became extinct.

However, in the words of Simberloff (2003), rapid responses to recent invaders will often 'resemble a blunderbuss attack rather than a surgical strike'. He notes, for example, that a string of successful eradications of small populations of weeds such as pampas grass (*Cortaderia selloana*) and ragwort (*Senecio jacobaea*) on New Zealand's offshore islands (Timmins & Braithwaite, 2002) were effective because of early action using brute-force methods. Similarly, the white-spotted tussock moth (*Orygia thyellina*), discovered in a suburban region of Auckland, New Zealand, was eradicated (at a cost of US\$5 million) using *Bacillus thuringiensis* spray (Clearwater, 2001). The only population biological information to hand was that females attracted males by pheromone, knowledge that was used to trap males and determine areas that needed respraying. Eradication of a recently established species known to be invasive elsewhere usually cannot and should not wait for new population studies to be performed.

Once invaders have established and spread through a new area and are determined to be pests, they are just another species at which the pest manager's armory must be directed.

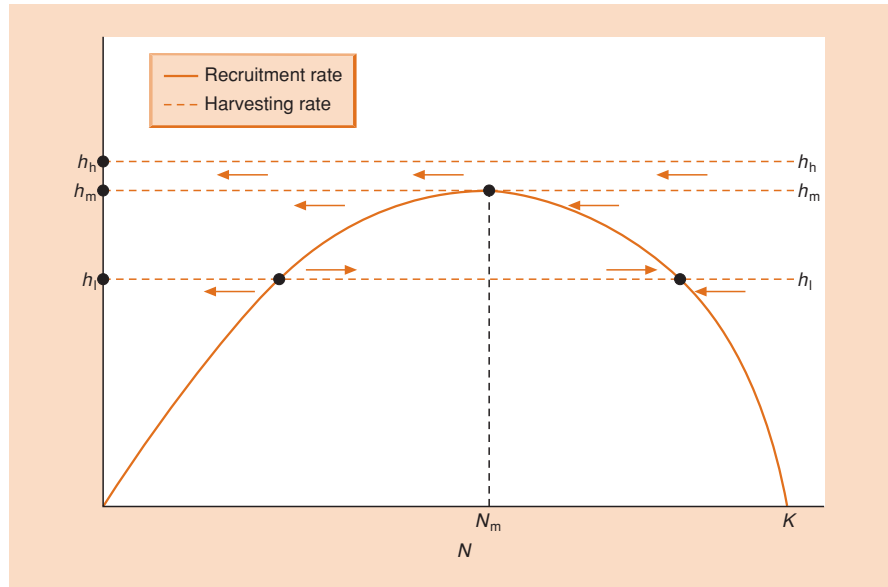
15.3 Harvest management

Harvesting of populations by people is clearly in the realm of predator-prey interactions and harvest management relies on the theory of predator-prey

harvesting aims to avoid over- and underexploitation

dynamics (see Chapters 10 and 14). When a natural population is exploited by culling or harvesting – whether this involves the removal of whales or fish from the sea, the capture of 'bushmeat' in the African savanna or the removal of timber from a forest – it is much easier to say what we want to avoid than precisely what we might wish to achieve. On the one hand, we want to avoid overexploitation, where too many individuals are removed and the population is driven into biological jeopardy, or economic insignificance or perhaps even to extinction. But harvest managers also want to avoid underexploitation, where far fewer individuals are removed than the population can bear, and a crop of food, for example, is produced which is smaller than necessary, threatening both the health of potential consumers and the livelihood of all those employed in the harvesting operation. However, as we shall see, the best position to occupy between these two extremes is not easy to determine, since it needs to combine considerations that are not only biological (the well-being of the exploited population) and economic (the profits being made

Figure 15.7 Fixed quota harvesting. The figure shows a single recruitment curve and three fixed quota harvesting curves: high quota (h_h), medium quota (h_m) and low quota (h_l). Arrows in the figure refer to changes to be expected in abundance under the influence of the harvesting rate to which the arrows are closest. ●, equilibria. At h_h the only ‘equilibrium’ is when the population is driven to extinction. At h_l there is a stable equilibrium at a relatively high density, and also an unstable breakpoint at a relatively low density. The MSY is obtained at h_m because it just touches the peak of the recruitment curve (at a density N_m): populations greater than N_m are reduced to N_m , but populations smaller than N_m are driven to extinction.



from the operation), but also social (local levels of employment and the maintenance of traditional lifestyles and human communities) (Hilborn & Walters, 1992; Milner-Gulland & Mace, 1998). We begin, though, with the biology.

15.3.1 Maximum sustainable yield

MSY: the peak of the net recruitment curve The first point to grasp about harvesting theory is that high yields are obtained from populations held below, often well below, the carrying capacity. This fundamental pattern is captured by the model population in Figure 15.7. There, the natural net recruitment (or net productivity) of the population is described by an n-shaped curve (see Section 5.4.2). Recruitment rate is low when there are few individuals and low when there is intense intraspecific competition. It is zero at the carrying capacity (K). The density giving the highest net recruitment rate depends on the exact form of intraspecific competition. This density is $K/2$ in the logistic equation (see Section 5.9) but, for example, is only slightly less than K in many large mammals (see Figure 5.10d). Always, though, the rate of net recruitment is highest at an ‘intermediate’ density, less than K .

Figure 15.7 also illustrates three possible harvesting ‘strategies’, although in each case there is a fixed harvesting rate, i.e. a fixed number of individuals removed during a given period of time, or ‘fixed quota’. When the harvesting and recruitment lines cross, the harvesting and recruitment rates are equal and opposite; the number removed per unit time by the harvester equals the number recruited per unit time by the population. Of particular interest is the harvesting rate h_m , the line that crosses (or, in fact, just touches) the recruitment rate curve at its peak. This is

the highest harvesting rate that the population can match with its own recruitment. It is known as the maximum sustainable yield (MSY), and as the name implies, it is the largest harvest that can be removed from the population on a regular and repeated (indeed indefinite) basis. It is equal to the maximum rate of recruitment, and it is obtained from the population by depressing it to the density at which the recruitment rate curve peaks.

The MSY concept is central to much of the theory and practice of harvesting. This makes the recognition of the following shortcomings in the concept all the more essential.

MSY has severe shortcomings . . .

- 1 By treating the population as a number of similar individuals, or as an undifferentiated biomass, it ignores all aspects of population structure such as size or age classes and their differential rates of growth, survival and reproduction. The alternatives that incorporate structure are considered below.
- 2 By being based on a single recruitment curve it treats the environment as unvarying.
- 3 In practice, it may be impossible to obtain a reliable estimate of the MSY.
- 4 Achieving an MSY is by no means the only, nor necessarily the best, criterion by which success in the management of a harvesting operation should be judged (see, for example, Section 15.3.9).

Despite all these difficulties, the MSY concept dominated resource management for many years in fisheries, forestry and wildlife exploitation. Prior to 1980, for example, there were 39 agencies for the management of marine fisheries, every

. . . but has been frequently used

one of which was required by its establishing convention to manage on the basis of an MSY objective (Clark, 1981). In many other areas, the MSY concept is still the guiding principle. Moreover, by assuming that MSYs are both desirable and attainable, a number of the basic principles of harvesting can be explained. Therefore, we begin here by exploring what can be learnt from analyses based on the MSY, but then look more deeply at management strategies for exploited populations by examining the various shortcomings of MSY in more detail.

15.3.2 Simple MSY models of harvesting: fixed quotas

fixed-quota harvesting is extremely risky . . .

The MSY density (N_m) is an equilibrium (gains = losses), but when harvesting is based on the removal of a fixed quota, as it is in Figure 15.7, N_m is a very fragile equilibrium. If the density exceeds the MSY density, then h_m exceeds the recruitment rate and the population declines towards N_m . This, in itself, is satisfactory. But if, by chance, the density is even slightly less than N_m , then h_m will once again exceed the recruitment rate. Density will then decline even further, and if a fixed quota at the MSY level is maintained, the population will decline until it is extinct. Furthermore, if the MSY is even slightly overestimated, the harvesting rate will always exceed the recruitment rate (h_h in Figure 15.7). Extinction will then follow, whatever the initial density. In short, a fixed quota at the MSY level might be desirable and reasonable in a wholly predictable world about which we had perfect knowledge. But in the real world of fluctuating environments and imperfect data sets, these fixed quotas are open invitations to disaster.

. . . whose dangers are illustrated by the Peruvian anchovy fishery

Nevertheless, a fixed-quota strategy has frequently been used. On a specified day in the year, the fishery (or hunting season) is opened and the cumulative catch logged. Then, when the quota (estimated MSY) has been taken, the fishery is closed for the rest of the year. An example of the use of fixed quotas is provided by the Peruvian anchovy (*Engraulis ringens*) fishery (Figure 15.8). From 1960 to 1972 this was the world's largest single fishery, and it constituted a major sector of the Peruvian economy. Fisheries experts advised that the MSY was around 10 million tonnes annually, and catches were limited accordingly. But the fishing capacity of the fleet expanded, and in 1972 the catch crashed. Overfishing seems at least to have been a major cause of the collapse, although its effects were compounded with the influences of profound climatic fluctuations. A moratorium on fishing would have been an ecologically sensible step, but this was not politically feasible: 20,000 people were dependent on the anchovy industry for employment. The stock took more than 20 years to recover (Figure 15.8).

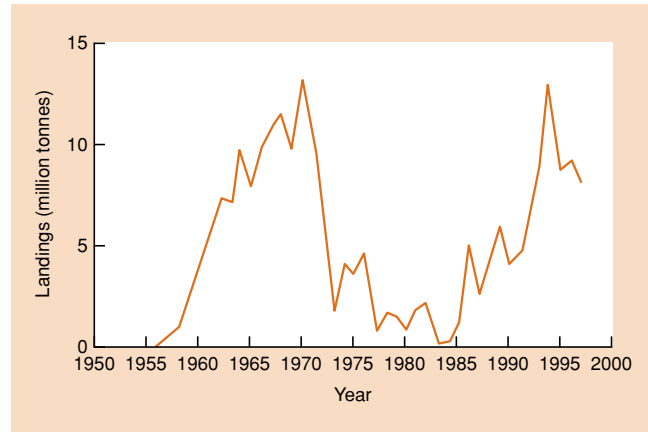


Figure 15.8 Landings of the Peruvian anchovy since 1950. (After Jennings *et al.*, 2001; data from FAO, 1995, 1998.)

15.3.3 A safer alternative: fixed harvesting effort

The risk associated with fixed quotas can be reduced if instead there is regulation of the harvesting effort. The yield from a harvest (H) can be thought of, simply, as being dependent on three things:

$$H = qEN. \quad (15.1)$$

Yield, H , increases with the size of the harvested population, N ; it increases with the level of harvesting effort, E (e.g. the number of 'trawler-days' in a fishery or the number of 'gun-days' with a hunted population); and it increases with harvesting efficiency, q . On the assumption that this efficiency remains constant, Figure 15.9a depicts an exploited population subjected to three potential harvesting strategies differing in harvesting effort. Figure 15.9b then illustrates the overall relationship to be expected, in a simple case like this, between effort and average yield: there is an apparently 'optimum' effort giving rise to the MSY, E_m , with efforts both greater and less than this giving rise to smaller yields.

regulating harvesting effort is less risky – but leads to a more variable catch

Adopting E_m is a much safer strategy than fixing an MSY quota. Now, in contrast to Figure 15.7, if density drops below N_m (Figure 15.9a), recruitment exceeds the harvesting rate and the population recovers. In fact, there needs to be a considerable overestimate of E_m before the population is driven to extinction (E_0 in Figure 15.9a). However, because there is a fixed effort, the yield varies with population size. In particular, the yield will be less than the MSY whenever the population size, as a result of natural fluctuations, drops below N_m . The appropriate reaction would be to reduce effort slightly or at least hold it steady whilst the population recovers. But an understandable (albeit misguided) reaction

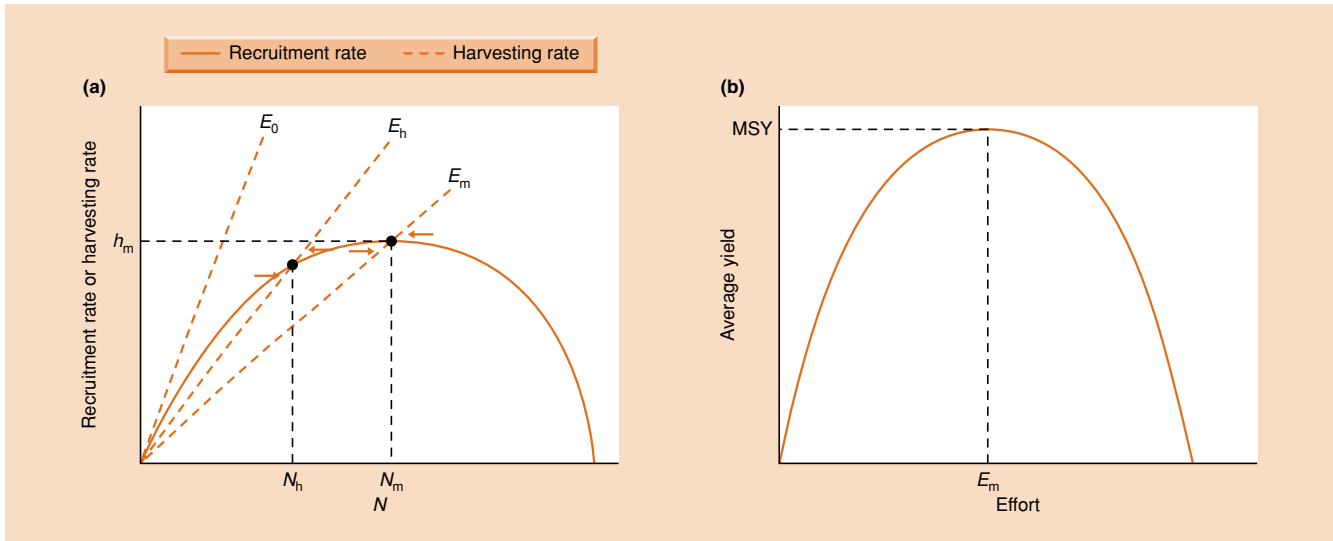


Figure 15.9 Fixed effort harvesting. (a) Curves, arrows and dots are as in Figure 15.7. The maximum sustainable yield (MSY) is obtained with an effort of E_m , leading to a stable equilibrium at a density of N_m with a yield of h_m . At a somewhat higher effort (E_h), the equilibrium density and the yield are both lower than with E_m but the equilibrium is still stable. Only at a much higher effort (E_0) is the population driven to extinction. (b) The overall relationship between the level of the fixed effort and average yield.

might be to compensate by increasing the effort. This, however, might depress population size further (E_h in Figure 15.9a); and it is therefore easy to imagine the population being driven to extinction as very gradual increases in effort chase an ever-diminishing yield.

There are many examples of harvests being managed by legislative regulation of effort, and this occurs in spite of the fact that effort usually defies precise measurement and control. For instance, issuing a number of gun licenses leaves the accuracy of the hunters uncontrolled; and regulating the size and composition of a fishing fleet leaves the weather to chance. Nevertheless, the harvesting of mule deer, pronghorn antelope and elk in Colorado was controlled by issuing a limited but varying number of hunting permits (Pojar, 1981). In the management of the important Pacific halibut stock, effort was limited by seasonal closures and sanctuary zones – although a heavy investment in fishery protection vessels was needed to make this work (Pitcher & Hart, 1982).

15.3.4 Other MSY approaches: harvesting a fixed proportion or allowing constant escapement

other MSY approaches: ... harvesting a constant proportion ...

Two further management strategies are based on the simple idea of availability of a surplus yield. First, a constant proportion of the population can be harvested (this is equivalent to fixing a hunting mortality rate and should

have the same effect as harvesting at constant effort) (Milner-Gulland & Mace, 1998). Thus, in the Northwest Territories of Canada, 3–5% of the caribou and muskox populations can be killed each year (Gunn, 1998), a strategy that involves the expense of preharvest censuses so that numbers to be harvested can be calculated.

Another strategy leaves a fixed number of breeding individuals at the end of each hunting season (constant escapement), an approach that involves the even greater expense of continuous monitoring through the hunting season. Constant escapement is a particularly safe option because it rules out the accidental removal of all the breeding individuals before breeding has occurred. Constant escapement is particularly useful for annual species because they lack the buffer provided by immature individuals in longer lived species (Milner-Gulland & Mace, 1998). The Falkland Islands government uses a constant escapement strategy for the annual *Loligo* squid. Stock sizes are assessed weekly from mid-season onwards and the fishery is closed when the ratio of stocks in the presence and absence of fishing falls to 0.3–0.4. After 10 years of this management regime the squid fishery shows good signs of sustainability (Figure 15.10).

Stephens *et al.* (2002) used simulation models to compare the outcomes for a population of alpine marmots (*Marmota marmota*) of fixed-quota, fixed-effort and threshold harvesting. In the latter case,

... or leaving a constant 'escapement' of breeding individuals

constant escapement seems to work best for alpine marmot hunting

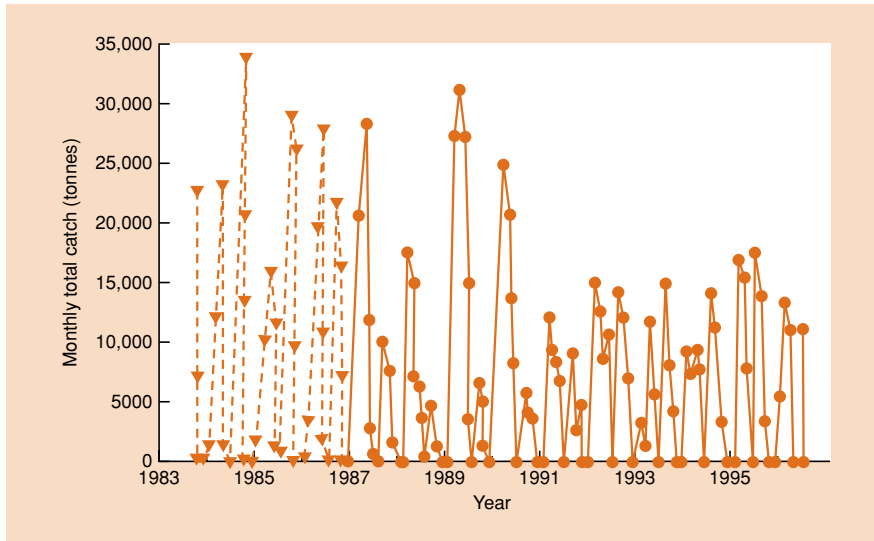


Figure 15.10 Monthly *Loligo* squid catches by licensed vessels in the Falkland Islands where a constant escapement management strategy is used. Note that there are two fishing seasons each year (February–May and August–October). The dotted lines (1984–86) represent estimated rather than actual catches. (After des Clers, 1998.)

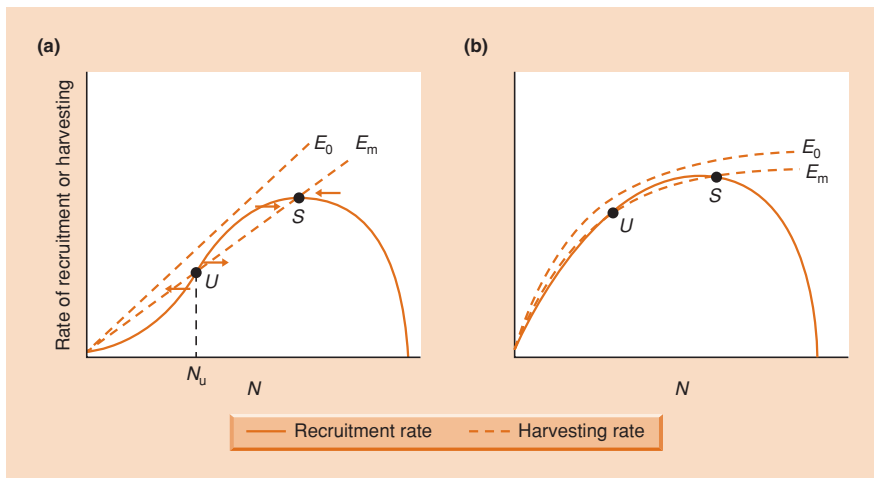


Figure 15.11 Multiple equilibria in harvesting. (a) When recruitment rate is particularly low at low densities, the harvesting effort giving the MSY (E_m) has not only a stable equilibrium (S) but also an unstable breakpoint (U) at a density below which the population declines to extinction. The population can also be driven to extinction by harvesting efforts (E_0) not much greater than E_m . (b) When harvesting efficiency declines at high densities, comments similar to those in (a) are appropriate.

harvesting only occurred during years in which the population exceeded a given threshold and exploitation continued until that threshold was reached (essentially a constant escapement approach). These social mammals are hunted in parts of Europe but the modeling was performed using extensive data available from a nonhunted population. They found that threshold harvesting provided the highest mean yields coupled with an acceptably low extinction risk. However, the introduction of error, associated with less frequent censuses (3-yearly rather than yearly), led to higher variance in yields and a much increased extinction probability (Stephens *et al.*, 2002). This emphasizes the importance of frequent censuses for constant escapement strategies to succeed.

15.3.5 Instability of harvested populations: multiple equilibria

Even with regulation of effort, harvesting near the MSY level may be courting disaster. The recruitment rate may be particularly low in the smallest populations (a pattern known as *depensation*; Figure 15.11a); for instance, the recruitment of young salmon is low at low densities because of intense predation from larger fish, and the recruitment of young whales may be low at low densities simply because of the reduced chances of males and females meeting to mate. However, depensation is apparently quite rare; Myers *et al.* (1995) detected it in only three of 128 fish stock

the problem of 'depensation'

data sets with 15 or more years available for analysis. Alternatively, harvesting efficiency may increase in small populations (Figure 15.11b). For instance, many clupeids (sardines, anchovies, herring) are especially prone to capture at low densities, because they form a small number of large schools that follow stereotyped migratory paths that the trawlers can intercept. With either depensation or higher harvesting efficiency at low density, small overestimates of E_m are liable to lead to overexploitation or even eventual extinction.

harvesting operations with multiple equilibria are susceptible to dramatic irreversible crashes

Even more important, however, is the fact that these interactions may have crucial ‘multiple equilibria’ (see Section 10.6). Note the two points where the harvesting line crosses the recruitment curve in Figure 15.11a. The point S is a stable equilibrium but the point U is an unstable ‘breakpoint’.

If the population is driven slightly below the MSY density, or even to a level slightly above N_U , a breakpoint, it returns to the MSY density (Figure 15.11a). But a marginally increased depression in density, to a level slightly below N_U , perhaps resulting from only a very small increase in effort, would make the harvesting rate greater than the recruitment rate. The population would be *en route* to extinction. Moreover, once the population is on this slippery slope, much more than a marginal reduction in effort is required to reverse the process. This is the crucial, practical point about multiple equilibria: a very slight change in behavior can lead to a wholly disproportionate change in outcome as the point of attraction in the system shifts from one stable state to another. Drastic changes in stock abundance can result from only small changes in harvesting strategy or small changes in the environment.

15.3.6 Instability of harvested populations: environmental fluctuations

It is tempting to attribute all fisheries’ collapses simply to overfishing and human greed. Doing so, however, would be an unhelpful oversimplification. There is no doubt that fishing pressure often exerts a great strain on the ability of natural populations to sustain levels of recruitment that counteract overall rates of loss. But the immediate cause of a collapse – in 1 year rather than any other – is often the occurrence of unusually unfavorable environmental conditions. Moreover, when this is the case, the population is more likely to recover (once conditions have returned to a more favorable state) than it would be if the crash was the result of overfishing alone.

the anchoveta and the El Niño

The Peruvian anchovy (see Figure 15.8), prior to its major collapse from 1972 to 1973, had already suffered a dip in the upward rise in catches in

the mid-1960s as a result of an ‘El Niño event’: the incursion of

warm tropical water from the north severely reducing ocean upwelling, and hence productivity, within the cold Peruvian current coming from the south (see Section 2.4.1). By 1973, however, because fishing intensity had so greatly increased, the effects of a subsequent El Niño event were much more severe. Moreover, whilst the fishery showed some signs of recovery from 1973 to 1982, in spite of largely unabated fishing pressure, a further collapse occurred in 1983 associated with yet another El Niño event. Clearly, it is unlikely that the consequences of these natural perturbations to the usual patterns of current flow would have been so severe if the anchovy had not been exploited or had been only lightly fished. It is equally clear, though, that the history of the Peruvian anchovy fishery cannot be understood properly in terms simply of fishing, as opposed to natural events.

The three Norwegian and Icelandic herring fisheries also collapsed in the early 1970s and had certainly been subjected to increasing fishing intensities prior to that. Once again, however, an oceanic anomaly is implicated (Beverton, 1993). In the mid-1960s, a mass of cold, low-salinity water from the Arctic Basin formed north of Iceland. It drifted south until it became entrained in the Gulf Stream several years later, and then moved north again – although well to the east of its southward track. It eventually disappeared off Norway in 1982 (Figure 15.12a). Data for the number of ‘recruits per spawner’, essentially the birth rate, are illustrated in Figure 15.12b for the Norwegian springspawning and the Icelandic spring- and summer-spawning herring between 1947 and 1990, in terms of the difference each year between that year’s value and the overall average. Also illustrated are the corresponding yearly temperature differentials in the Norwegian Sea, reflecting the southward and northward passage of the anomalous cold water body. There was a good correspondence between the cold water and poor recruitment in both the Icelandic and Norwegian stocks in the late 1960s and in the Norwegian stocks in 1979–81, the Icelandic stocks being then extinct (spring spawners) or too far west. It seems likely that the anomalous cold water led to unusually low recruitment, which was strongly instrumental in the crashes experienced by each of these fisheries.

herring and cold water

This cannot, however, account for all the details in Figure 15.12b – especially the succession of poor recruitment years in the Norwegian stocks in the 1980s. For this, a more complex explanation is required, probably involving other species of fish and perhaps alternative stable states (Beverton, 1993). None the less, it remains clear that whilst the dangers of overfishing should not be denied, these must be seen within the context of marked and often unpredictable natural variations. Given the likely effects of environmental conditions on the vital rates of harvested populations, a reliance on models with constant vital rates is even more risky. Engen *et al.* (1997) argue that the best harvesting strategies for such highly variable populations involve constant escapement (see Section 15.3.4).

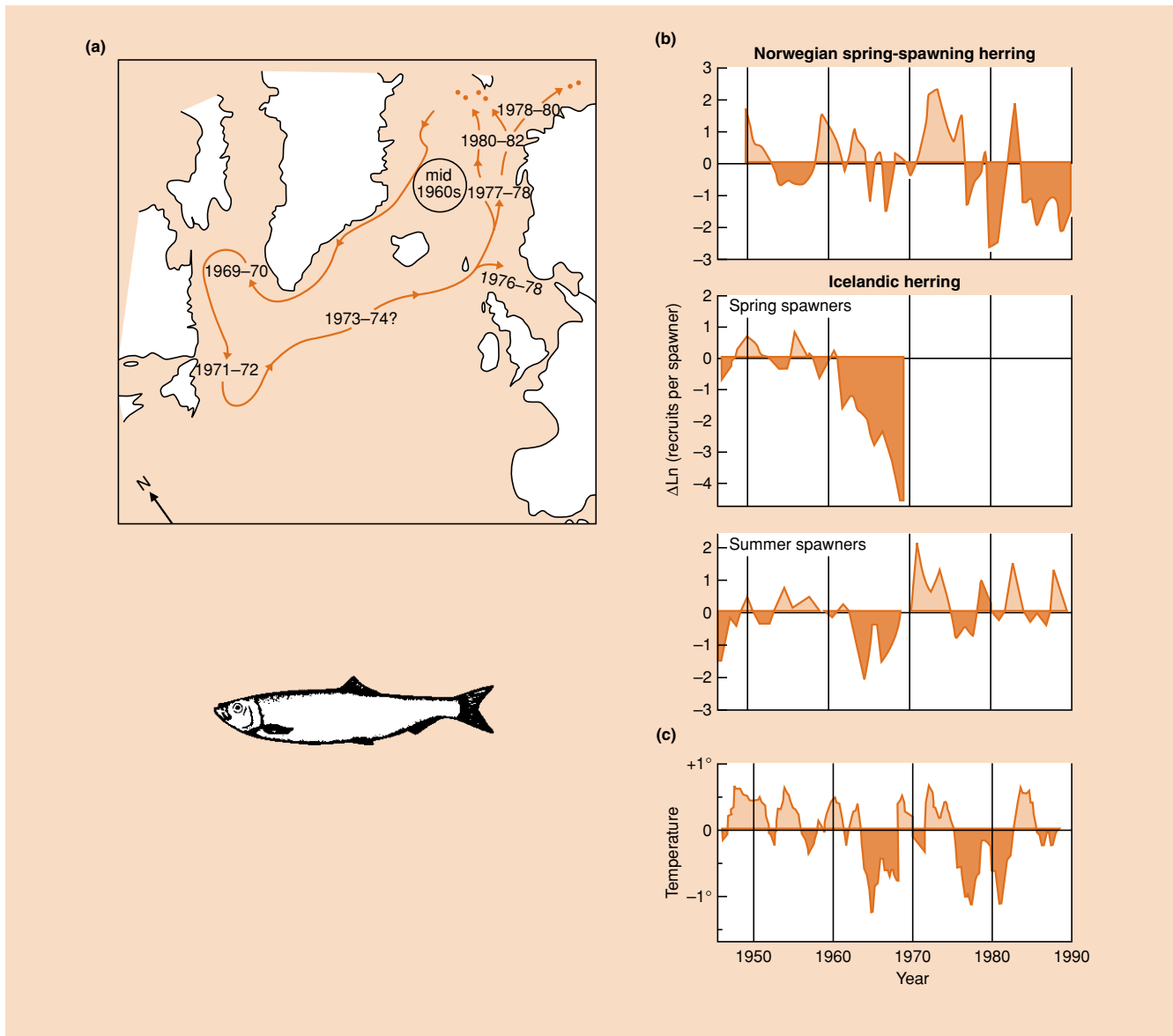


Figure 15.12 (a) The track of a large mass of cold, low-salinity water in the 1960s and 1970s, showing its presence in the Norwegian Sea both in the mid-1960s and the period 1977–82. (b) Annual differentials between overall averages and \ln (recruits per spawner) for three herring stocks in the Norwegian Sea, and (c) the temperature in the Norwegian Sea. The Icelandic spring-spawning stock never recovered from its collapse in the early 1970s, preceded by low recruitment in the 1960s. (After Beverton, 1993.)

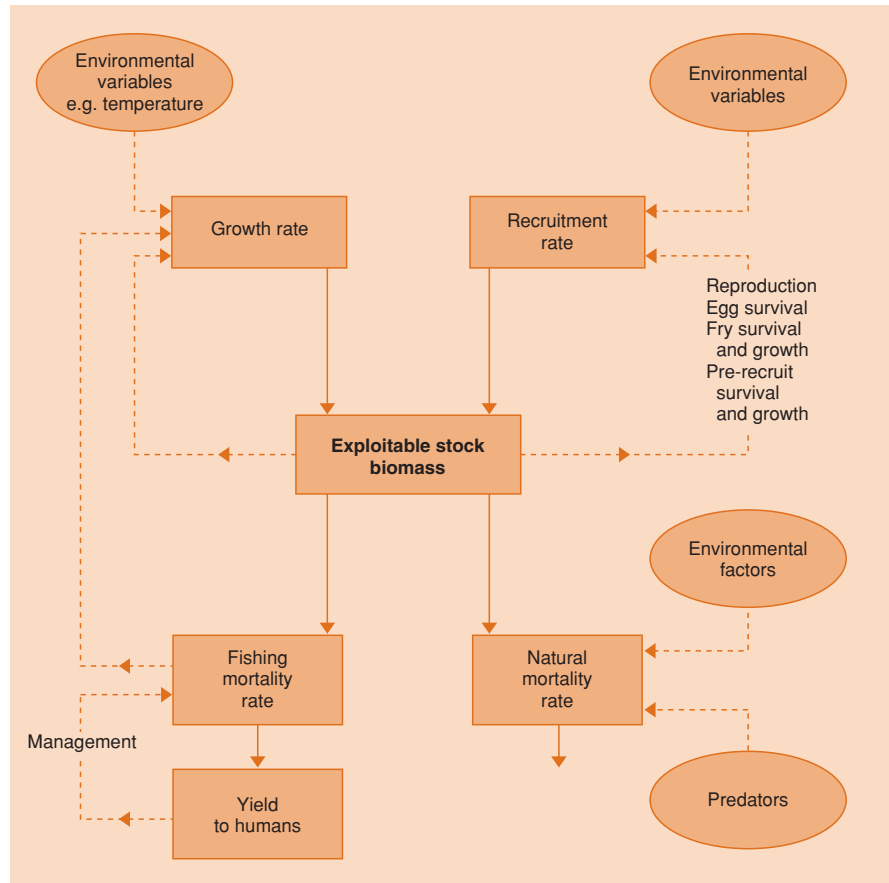
15.3.7 Recognizing structure in harvested populations: dynamic pool models

‘dynamic pool’ models recognize population structure

The simple models of harvesting that have been described so far are known as ‘surplus yield’ models. They are useful as a means of establishing some basic principles (like MSY), and they are good for investigating the possible consequences of different types of harvesting strategy.

But they ignore population structure, and this is a bad fault for two reasons. The first is that ‘recruitment’ is, in practice, a complex process incorporating adult survival, adult fecundity, juvenile survival, juvenile growth, and so on, each of which may respond in its own way to changes in density and harvesting strategy. The second reason is that most harvesting practices are primarily interested in only a portion of the harvested population (e.g. mature trees, or fish that are large enough to be saleable). The approach that attempts to take these complications into

Figure 15.13 The dynamic pool approach to fishery harvesting and management, illustrated as a flow diagram. There are four main ‘submodels’: the growth rate of individuals and the recruitment rate into the population (which add to the exploitable biomass), and the natural mortality rate and the fishing mortality rate (which deplete the exploitable biomass). Solid lines and arrows refer to changes in biomass under the influence of these submodels. Dashed lines and arrows refer to influences either of one submodel on another, or of the level of biomass on a submodel or of environmental factors on a submodel. Each of the submodels can itself be broken down into more complex and realistic systems. Yield to humans is estimated under various regimes characterized by particular values inserted into the submodels. These values may be derived theoretically (in which case they are ‘assumptions’) or from field data. (After Pitcher & Hart, 1982.)



account involves the construction of what are called ‘dynamic pool’ models.

The general structure of a dynamic pool model is illustrated in Figure 15.13. The submodels (recruitment rate, growth rate, natural mortality rate and fishing rate of the exploited stock) combine to determine the exploitable biomass of the stock and the way this translates into a yield to the fishing community. In contrast to the surplus yield models, this biomass yield depends not only on the number of individuals caught but also on their size (past growth); whilst the quantity of exploitable (i.e. catchable) biomass depends not just on ‘net recruitment’ but on an explicit combination of natural mortality, harvesting mortality, individual growth and recruitment into catchable age classes.

There are many variants on the general theme (e.g. the submodels can be dealt with separately in each of the age classes and submodels can incorporate as much or as little information as is available or desirable). In all cases, though, the basic approach is the same. Available information (both theoretical and empirical) is incorporated into a form that reflects the dynamics of the structured population. This then allows the yield and the response of the population to different harvesting strategies to be estimated.

This in turn should allow a recommendation to the stock-manager to be formulated. The crucial point is that in the case of the dynamic pool approach, a harvesting strategy can include not only a harvesting intensity, but also a decision as to how effort should be partitioned amongst the various age classes.

A classic example of a dynamic pool model in action concerned the Arcto-Norwegian cod fishery, the most northerly of the Atlantic stocks (Garrod & Jones, 1974). The age class structure

dynamic pool models can lead to valuable recommendations . . .

of the late 1960s was used to predict the medium-term effects on yield of different fishing intensities and different mesh sizes in the trawl. Some of the results are shown in Figure 15.14. The temporary peak after 5 or so years is a result of the very large 1969 year-class working through the population. Overall, however, it is clear that the best longer term prospects were predicted for a low fishing intensity and a large mesh size. Both of these give the fish more opportunity to grow (and reproduce) before they are caught, which is important because yield is measured in biomass, not simply in numbers. Higher fishing intensities and mesh sizes of 130 mm were predicted to lead to overexploitation of the stock.

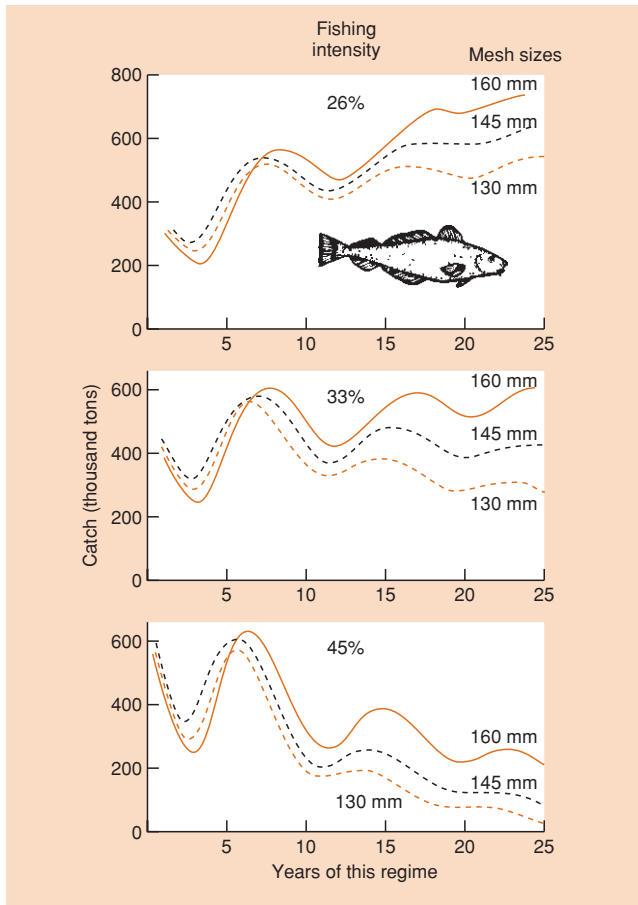


Figure 15.14 Garrod and Jones' (1974) predictions for the Arctic cod stock under three fishing intensities and with three different mesh sizes. (After Pitcher & Hart, 1982.)

... but these may still be ignored

Sadly, Garrod and Jones' recommendations were ignored by those with the power to determine fishing strategies. Mesh sizes were not increased until 1979, and then only from 120 to 125 mm. Fishing intensity never dropped below 45% and catches of 900,000 tonnes were taken in the late 1970s. Not surprisingly perhaps, surveys late in 1980 showed that these and other North Atlantic cod stocks were very seriously depleted as a result of overfishing. North Sea cod reach sexual maturity around the age of 4 years, but the species has been so heavily exploited that some 1 year olds are now harvested and 2 year olds are almost fully exploited, leaving only 4% of 1 year olds to survive to age 4 (Cook *et al.*, 1997).

Rattans (climbing spiny palms whose stems are used for weaving and furniture making in Southeast Asia) are threatened with overexploitation in a similar way, with harvesters cutting stems too young and reducing their ability to resprout (MacKinnon, 1998).

15.3.8 Objectives for managing harvestable resources

If we treat the Garrod and Jones example as typical, then we might conclude that the biologist proposes – but the manager disposes. This is therefore an appropriate point at which to reconsider not only the objectives of harvesting programs, but also the criteria by which successful management should be judged and the role of ecologists in management overall. As Hilborn and Walters (1992) have pointed out, there are three alternative attitudes that ecologists can take, each of which has been popular but only one of which is sensible. Indeed, these are increasingly important considerations that apply not just to fisheries management but to every entry of ecologists into the public arena.

The first is to claim that ecological interactions are too complex, and our understanding and our data too poor, for pronouncements of any kind to be made (for fear of being wrong). The problem with this is that if ecologists choose to remain silent because of some heightened sensitivity to the difficulties, there will always be some other, probably less qualified 'expert' ready to step in with straightforward, not to say glib, answers to probably inappropriate questions.

The second possibility is for ecologists to concentrate exclusively on ecology and arrive at a recommendation designed to satisfy purely ecological criteria. Any modification by managers or politicians of this recommendation is then ascribed to ignorance, inhumanity, political corruption or some other sin or human foible. The problem with this attitude is that it is simply unrealistic in any human activity to ignore social and economic factors.

The third alternative, then, is for ecologists to make ecological assessments that are as accurate and realistic as possible, but to assume that these will

be incorporated with a broader range of factors when management decisions are made. Moreover, these assessments should themselves take account of the fact that the ecological interactions they address include humans as one of the interacting species, and humans are subject to social and economic forces. Finally, since ecological, economic and social criteria must be set alongside one another, choosing a single, 'best' option is likely to be seen by some involved in the decision as an opinion based on the proponent's particular set of values. It follows that a single recommendation is, in practice, far less useful in this discourse than laying out a series of possible plans of action with their associated consequences.

In the present context, therefore, we develop this third alternative by first looking beyond MSY to criteria that incorporate risk, economics, social consequences, and so on (Hilborn & Waters, 1992). We then briefly examine the means by which crucial parameters and variables are estimated in natural populations, since these, by determining the quality of available information, determine the degree of confidence with which recommendations can be made.

three attitudes for ecologists towards managers in the real world . . .

. . . but only one of them is sensible

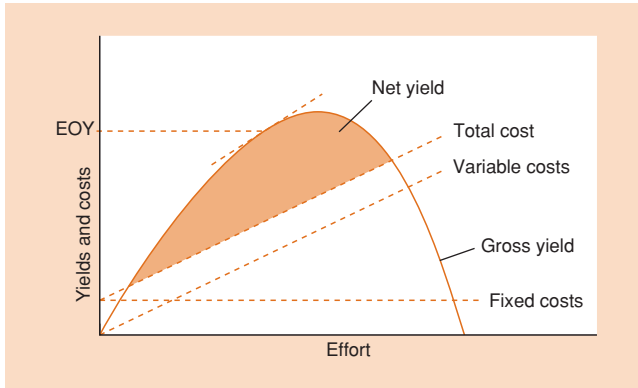


Figure 15.15 The economically optimum yield (EOY), that which maximizes ‘profit’, is obtained to the left of the peak of the yield-against-effort curve, where the difference between gross yield and total cost (fixed costs plus variable costs) is greatest. At this point, the gross yield and total cost lines have the same slope. (After Hilborn & Walters, 1992.)

15.3.9 Economic and social factors

the economically optimum yield – typically less than the MSY

Perhaps the most obvious shortcoming of a purely ecological approach is its failure to recognize that the exploitation of a natural resource is usually a business enterprise, in which the value of

the harvest must be set against the costs of obtaining that harvest. Even if we distance ourselves from any preoccupation with ‘profit’, it makes no sense to struggle to obtain the last few tonnes of an MSY if the money spent in doing so could be much more effectively invested in some other means of food production. The basic idea is illustrated in Figure 15.15. We seek to maximize not total yield but net value – the difference between the gross value of the harvest and the sum of the fixed costs (interest payments on ships or factories, insurance, etc.) and the variable costs, which increase with harvesting effort (fuel, crew’s expenses, etc.). This immediately suggests that the economically optimum yield (EOY) is less than the MSY, and is obtained through a smaller effort or quota. However, the difference between the EOY and the MSY is least in enterprises where most costs are fixed (the ‘total cost’ line is virtually flat). This is especially the case in high investment, highly technological operations such as deep-sea fisheries, which are therefore most prone to overfishing even with management aimed at economic optima.

discounting: liquidating stocks, or leaving them to grow?

A second important economic consideration concerns ‘discounting’. This refers to the fact that in economic terms, each bird in the hand now (or each fish in the hold) is worth more than an equivalent bird or fish some time in

the future. The reason is basically that the value of the current catch can be placed in the bank to accrue interest, so that its total value increases. In fact, a commonly used discount rate for natural resources is 10% per annum (90 fish now are as valuable as 100 fish in 1 year’s time) despite the fact that the difference between the interest rates in the banks and the rate of inflation is usually only 2–5%. The economists’ justification for this is a desire to incorporate ‘risk’. A fish caught now has already been caught; one still in the water might or might not be caught – a bird in the hand really is worth two in the bush.

On the other hand, the caught fish is dead, whereas the fish still in the water can grow and breed (although it may also die). In a very real sense, therefore, each uncaught fish will be worth more than ‘one fish’ in the future. In particular, if the stock left in the water grows faster than the discount rate, as is commonly the case, then a fish put on deposit in the bank is not so sound an investment as a fish left on deposit in the sea. Nevertheless, even in cases like this, discounting provides an economic argument for taking larger harvests from a stock than would otherwise be desirable.

Moreover, in cases where the stock is less productive than the discount rate – for example, many whales and a number of long-lived fish – it seems to make sense, in purely economic terms, not only to overfish the stock, but actually to catch every fish (‘liquidate the stock’). The reasons for not doing so are partly ethical – it would clearly be ecologically short sighted and a disdainful way of treating the hungry mouths to be fed in the future. But there are also practical reasons: jobs must be found for those previously employed in the fishery (or their families otherwise provided for), alternative sources of food must be found, and so on. This emphasizes, first, that a ‘new economics’ must be forged in which value is assigned not only to things that can be bought and sold – like fish and boats – but also to more abstract entities, like the continued existence of whales or other ‘flagship species’ (Hughey *et al.*, 2002). It also stresses the danger of an economic perspective that is too narrowly focused. The profitability of a fishery cannot sensibly be isolated from the implications that the management of the fishery has in a wider sphere.

social repercussions

‘Social’ factors enter in two rather separate ways into plans for the management of natural resources. First, practical politics might dictate, for instance, that a large fleet of small, individually inefficient boats is maintained in an area where there are no alternative means of employment. In addition, though, and of much more widespread importance, it is necessary for management plans to take full account of the way fishermen and harvesters will behave and respond to changing circumstances, rather than assuming that they will simply conform to the requirements for achieving either ecological or economic optima. Harvesting involves a predator–prey interaction: it makes no sense to base plans on the dynamics of the prey alone whilst simply ignoring those of the predator (us!).

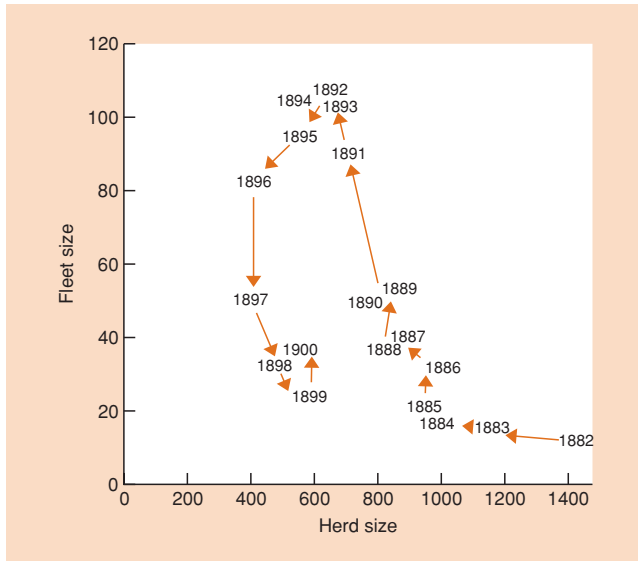


Figure 15.16 The fleet size of the North Pacific fur seal fishery (predators) responded to the size of the seal herd (prey) between 1882 and 1900 by exhibiting an anticlockwise predator–prey spiral. (After Hilborn & Walters, 1992; from data of Wilen, 1976, unpublished observations.)

harvester as predator: human behavior

The idea of the harvester as predator is reinforced in Figure 15.16, which shows a classic anticlockwise predator–prey spiral (see Chapter 10) for the North Pacific fur seal fishery in the last years of the 19th century. The figure illustrates a numerical response on the part of the predator – extra vessels enter the fleet when the stock is abundant, but leave when it is poor. But the figure also illustrates the inevitable time lag in this response. Thus, whatever a modeler or manager might propose, there is unlikely ever to be some perfect match, at an equilibrium, between stock size and effort. Moreover, whilst the sealers in the figure left

the fishery as quickly as they had entered it, this is by no means a general rule. The sealers were able to switch to fishing for halibut, but such switches are often not easy to achieve, especially where there has been heavy investment in equipment or long-standing traditions are involved. As Hilborn and Walters (1992) put it, ‘Principle: the hardest thing to do in fisheries management is reduce fishing pressure’.

Switching is one aspect of a harvester’s predatory behavior – its functional response (see Chapter 10). Harvesters will also generally ‘learn’ as there is an inevitable trend towards technological improvement. Even without this, harvesters usually improve their efficiency as they learn more about their stock – notwithstanding the assumptions of simple fixed-effort models.

15.3.10 Estimates from data: putting management into practice

The role of the ecologist in the management of a natural resource is in *stock assessment*: making quantitative predictions about the response of the biological population to alternative management choices and addressing questions like whether a given fishing intensity will lead to a decline in the size of the stock, whether nets of a given mesh size will allow the recruitment rate of a stock to recover, and so on. In the past, it has often been assumed that this can be done simply by careful monitoring. For example, as effort and yield increase in an expanding fishery, both are monitored, and the relationship between the two is plotted until it seems that the top of a curve like that in Figure 15.7 has been reached or just exceeded, identifying the MSY. This approach, however, is deeply flawed, as can be seen from Figure 15.17. In 1975, the International Commission for the Conservation of Atlantic Tunas (ICCAT) used the available data (1964–73) to plot the yield–effort relationship for the yellowfin tuna (*Thunnus albacares*) in the eastern Atlantic. They felt that they

monitoring effort and yield: the difficulties of ‘finding the top of the curve’

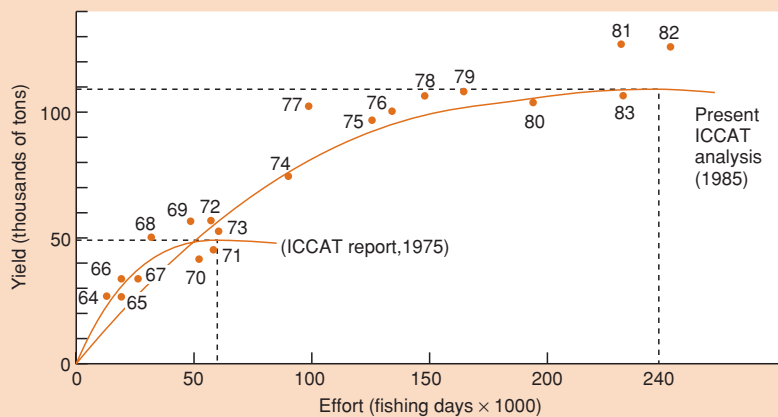


Figure 15.17 Estimated yield–effort relationships for the eastern Atlantic yellowfin tuna (*Thunnus albacares*) on the basis of the data for 1964–73 (ICCAT, 1975) and 1964–83 (ICCAT, 1985). (After Hunter *et al.*, 1986; Hilborn & Walters, 1992.)

had reached the top of the curve: a sustainable yield of around 50,000 tons (5.1×10^7 kg) and an optimum effort of about 60,000 fishing days. However, ICCAT were unable to prevent effort (and yield) rising further, and it soon became clear that the top of the curve had not been reached. A reanalysis using data up to 1983 suggested a sustainable yield of around 110,000 tons (1.1×10^8 kg) and an effort of 240,000 fishing days.

This illustrates what Hilborn and Walters (1992) describe as another principle: ‘You cannot determine the potential yield from a fish stock without overexploiting it’. At least part of the reason for this is the tendency, already noted, for the variability in yield to increase as an MSY is approached. Furthermore, if we also recall the previously described difficulty in reducing fishing pressure, it is clear that in practice, managers are likely to have to wrestle with the combined challenge of estimation difficulties, ecological relationships (here, between yield and predictability) and socio-economic factors (here, concerning the regulation and reduction in effort). We have moved a long way from the simple fixed-effort models of Section 15.3.3.

The practical difficulties of parameter estimation are further illustrated in Figure 15.18, which displays the time series for total catch, fishing effort and catch per unit effort (CPUE) between 1969 and 1982 for yellowfin tuna for the whole Atlantic Ocean. As effort increased, CPUE declined – presumably, a reflection of a diminishing stock of fish. On the other hand, the catch continued to rise over this period, suggesting that perhaps the stock was not yet being overfished (i.e. the MSY had not yet been reached). These, then, are the data, and they come in probably the most commonly available form – a so-called ‘one-way trip’ time series. But can they suggest an MSY and can they suggest the effort required to achieve that MSY? Certainly, methods exist for performing the necessary calculations, but these methods require assumptions to be made about the underlying dynamics of the population.

estimates from catch and effort data: applying the Schaefer model

The most frequently used assumption describes the dynamics of the stock biomass, B , by:

$$\frac{dB}{dt} = rB \left(1 - \frac{B}{K} \right) - H \tag{15.2}$$

(Schaefer, 1954), which is simply the logistic equation of Chapter 5 (intrinsic rate of increase, r , carrying capacity, K) with a harvesting rate incorporated. The latter may itself be given, following Equation 15.1 (see Section 15.3.3), by $H = qEB$, where q is harvesting efficiency and E the harvesting effort. By definition:

$$CPUE = H/E = qB. \tag{15.3}$$

Hence:

$$B = CPUE/q \tag{15.4}$$

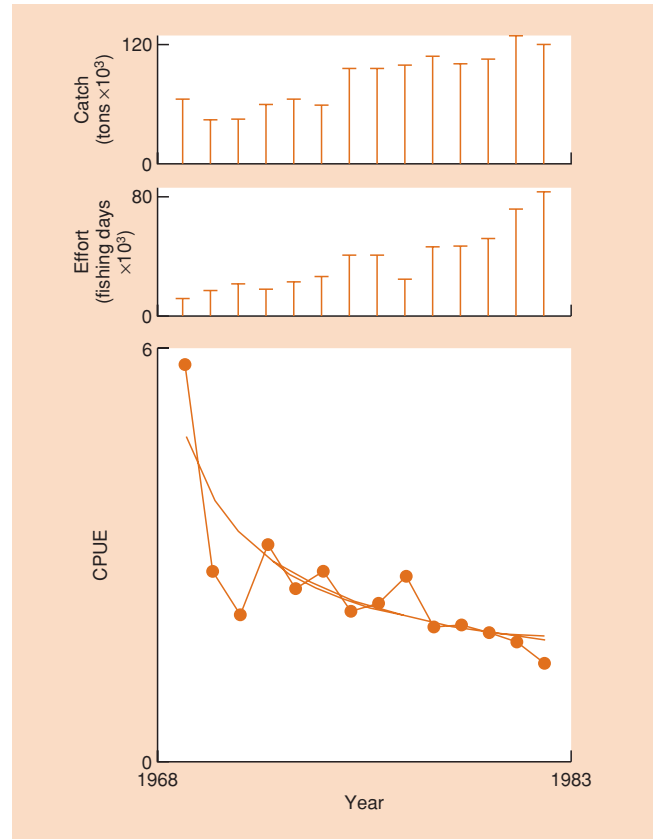


Figure 15.18 Changes in total catch, fishing effort and catch per unit effort (CPUE) between 1969 and 1982 for the yellowfin tuna (*Thussus albacares*) in the Atlantic Ocean. Also shown are three separate curves fitted to the CPUE time series by methods outlined in the text, the parameters of which are given in Table 15.4. (After Hilborn & Walters, 1992.)

and Equation 15.2 can be rewritten in terms of CPUE with either H or E as variables, and with r , q and K as parameters. For this model, the MSY is given by $rK/4$ and the effort required to achieve this by $r/2q$.

There are a number of methods of obtaining estimates of these parameters from field data, perhaps the best of which is the fitting of curves to time series (Hilborn & Walters, 1992).

time series analysis is best – but answers are still equivocal

However, when the time series is a one-way trip, as we have noted it often is, there is no unique ‘best’ set of parameter values. Table 15.4, for instance, shows the parameters for three separate curves fitted to the data in Figure 15.18, providing equally good fits (the same sum of squares), but with widely differing parameter values. There are, in effect, a large number of equally good alternative explanations for the data in Figure 15.18, in some of which, for example, the population has a low carrying capacity but a high intrinsic rate of increase and is being harvested

Table 15.4 Parameter estimates from three fits to the catch per unit effort (CPUE) time series for yellowfin tuna shown in Figure 15.18. r is the intrinsic rate of increase, K is the carrying capacity (equilibrium abundance in the absence of harvesting) and q is the harvesting efficiency. Effort is measured in fishing days; K and maximum sustainable yield (MSY) in tons. (After Hilborn & Walters, 1992.)

Fit number	r	$K (\times 1000)$	$q (\times 10^{-7})$	MSY ($\times 1000$)	Effort at MSY ($\times 1000$)	Sum of squares
1	0.18	2103	9.8	98	92	3.8
2	0.15	4000	4.5	148	167	3.8
3	0.13	8000	2.1	261	310	3.8

efficiently, whereas in others it has a high carrying capacity, a low rate of increase and is being harvested less efficiently. In the first case, the MSY had probably already been reached in 1980; in the second, catches could probably be doubled with impunity. Moreover, in each of these cases, the population is assumed to be behaving in conformity with Equation 15.2, which may itself be wide of the mark.

these uncertainties make ecologists all the more valuable

It is clear, therefore, even from this limited range of rather arbitrarily chosen examples, that there are immense limitations placed on stock assessments and management plans by inadequacies in both the available data and the means of analyzing them. This is not meant, though, to be a council of despair. Management decisions must be made, and the best possible stock assessments must form the basis – although not the sole basis – for these decisions. It is regrettable that we do not know more, but the problem would be compounded if we pretended that we did. Moreover, the ecological, economic and human behavioral analyses are important – as all analyses are – for identifying what we do *not* know, since, armed with this knowledge, we can set about obtaining whatever information is most useful. This has been formalized, in fact, in an ‘adaptive management’ approach, where, in an ‘actively adaptive’ strategy, a policy is sought which offers some balance between, on the one hand, probing for information (directed experimentation), and on the other, exercising caution about losses in short-term yield and long-term overfishing (Hilborn & Walters, 1992). Indeed, there is a strong argument that says that the inadequacies in data and theory make the need for ecologists all the more profound: who else can appreciate the uncertainties and provide appropriately enlightened interpretations?

‘dataless management’ where no estimates are available?

However, to be realistic, managing most marine fisheries to achieve optimum yields will be very difficult to achieve. There are generally too few researchers to do the work and, in many parts of the world, no researchers at all. In these situations, a precautionary approach to fisheries management might involve locking away a proportion of a coastal or coral community in marine protected areas (Hall, 1998). The term *dataless management* has been applied to situations where local villagers follow simple

prescriptions to make sustainability more likely. For example locals on the Pacific island of Vanuatu were provided with some simple principles of management for their trochus (*Tectus niloticus*) shellfishery (stocks should be harvested every 3 years and left unharvested in between) with an apparently successful outcome (Johannes, 1998).

15.4 The metapopulation perspective in management

A repeated theme in previous chapters has been the spatial patchiness upon which population interactions are often played out. Managers need to understand the implications of such heterogeneous landscape structure when making their decisions. Various approaches are available to improve our understanding of populations in complex landscapes and we consider two in the following sections. First, landscapes with different degrees of habitat loss and fragmentation can be artificially created at a scale appropriate to populations of interest and their behavior can then be assessed in carefully controlled experiments (see Section 15.4.1 – in the context of biological control of pests). Second, simple deterministic models can throw light on the factors that need to be taken into account when managing populations in a habitat patchwork (see Section 15.4.2 – in the context of creating protected areas for fisheries management). We also saw earlier (see Section 7.5.6 – in the context of a reserve patchwork for an endangered species) how stochastic simulation models can be used to compare management scenarios where subpopulations exist in a metapopulation.

15.4.1 Biological control in a fragmented landscape

We know that spatial heterogeneity can stabilize predator–prey interactions (e.g. Chapter 10). However, the dynamics of pests and their biological control agents may become destabilized, resulting in pest outbreaks, if habitat change occurs at a scale that

natural enemy success may depend on predation efficiency in a patchy habitat

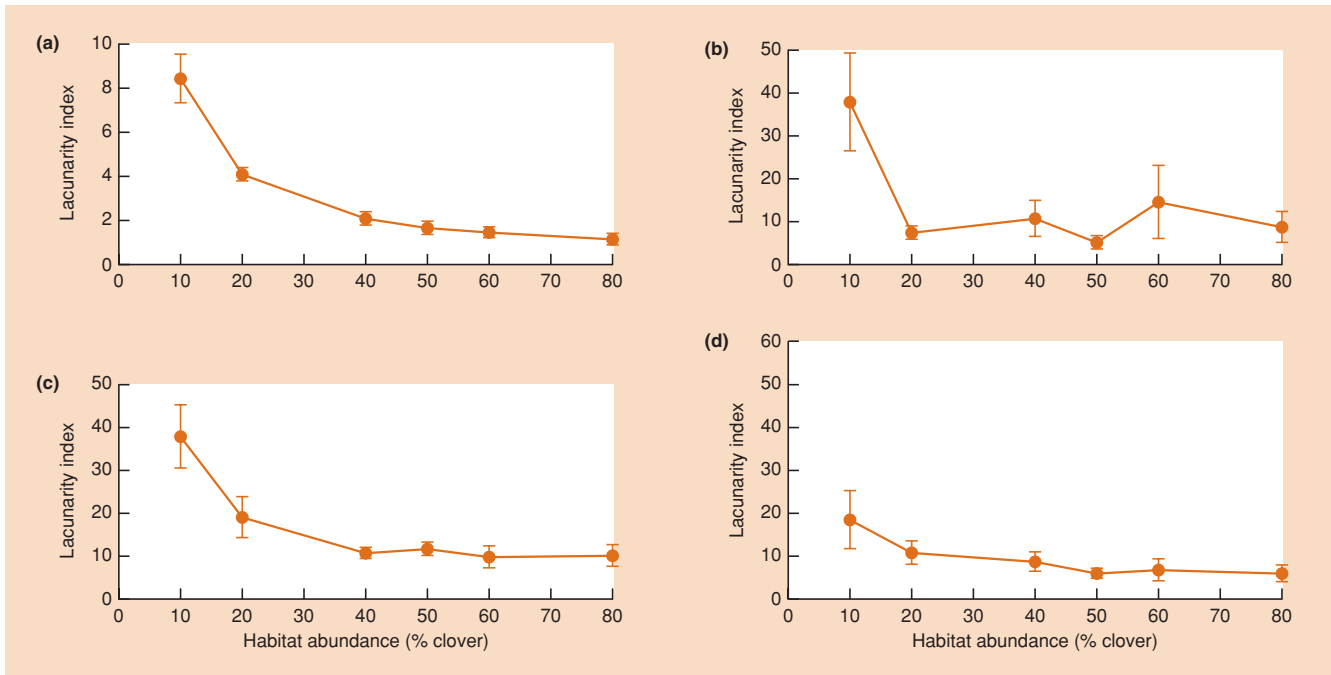


Figure 15.19 The distribution pattern (lacunarity index – a measure of aggregation) of (a) clover (i.e. habitat) and populations of (b) pest aphids, (c) an introduced ladybird beetle control agent (*Harmonia axyridis*) and (d) a native ladybird beetle (*Coleomegilla maculata*). In these experiments, clover plants were clumped together as opposed to being dispersed through the landscape. Error bars are ±1 SE. (After With *et al.*, 2002.)

interferes with the search behavior of a control agent (Kareiva, 1990).

With *et al.* (2002) created replicate landscapes (plots) of red clover (*Trifolium pratense*), each 16 × 16 m, that differed in terms of clover abundance (10, 20, 40, 50, 60 and 80% *T. pratense*). Their aim was to explore whether thresholds in landscape structure precipitate similar thresholds in the distribution of a pest aphid, *Acyrtosiphon pisum*, and to discover how landscape structure affects the search behavior of two ladybird beetle predators of aphids, one an introduced biocontrol agent, *Harmonia axyridis*, the other a native species, *Coleomegilla maculata*. Colonization by the aphids and beetles was by natural immigration to the outdoor plots.

Lacunarity is an index of aggregation derived from fractal geometry that quantifies the variability in the distribution of gap sizes (distances among clover patches in the landscape). The distribution of clover in the experimental landscapes showed a threshold at 20% habitat, indicating that gap sizes became greater and more variable below this level (Figure 15.19a). This threshold was mirrored by the aphids (Figure 15.19b) and was strongly tracked by the exotic control agent (*H. axyridis*) but not the native predator (*C. maculata*) (Figure 15.19c, d).

Although the native ladybird foraged more actively among stems within the clover cells, overall it was less mobile and moved less between clover cells in the landscape than the introduced

ladybird, which showed a greater tendency to fly (Table 15.5). With its greater mobility, the introduced species was more effective at tracking aphids when they occurred at low patch occupancy, a prerequisite for successful biological control (Murdoch & Briggs, 1996).

Findings such as these have implications both for the selection of effective biological control agents and for the design of agricultural systems, which may need to be managed to preserve habitat connectivity and thus enhance the efficiency of natural enemies and/or biological control agents (Barbosa, 1998).

15.4.2 Designing reserve networks for fisheries management

Over the last decade or so, coastal marine reserves or no-take zones have been promoted as a means of managing fisheries (e.g. Holland & Brazeal, 1996). This is another example where an understanding of landscape structure, and metapopulation dynamics, will be necessary to devise management strategies. Probably the most fundamental questions of reserve design are the fraction of coastline that should be set aside and the appropriate size (and number) of reserves needed in relation to the

fishery management using no-take zones: metapopulation considerations

<i>Scale and behavior measure</i>	<i>Introduced Harmonia axyridis</i>	<i>Native Coleomegilla maculata</i>
Within clover cells		
Stems visited per minute	0.80 ± 0.05	1.20 ± 0.07
Between clover cells		
Cells visited per minute	0.22 ± 0.07	0.10 ± 0.04
Primary mode of movement	Fly	Crawl
Plot-wide movement		
Mean step length (m)	1.90 ± 0.21	1.10 ± 0.04
Displacement ratio	0.49 ± 0.05	0.19 ± 0.03

Table 15.5 Search behavior of introduced and native ladybird beetles at different scales in experimental clover landscapes. Values are means ± 1 SE. Each 16 × 16 m plot contains 256 cells (each 1 m²); clover cells are those cells in which clover was present. For individual ladybirds that made at least five cell transitions, plot-wide movements were quantified in terms of mean step length and displacement ratio. Displacement ratio is net displacement (straight-line distance) divided by overall path length. (After With *et al.*, 2002.)

dispersal potential of the target species. Hastings and Botsford (2003) developed a simple deterministic model to answer these questions for a hypothetical species with characteristics that are most likely to benefit from no-take zones: one with sedentary adults and dispersing larvae. Their approach is based on the idea that altering the spacing and width of reserves changes the fraction of larvae that are retained within or exported from reserves (Figure 15.20). It is, of course, larval export that provides the basis for a sustainable yield from nonreserve areas.

The MSY problem can be stated as ‘fix the level of larval retention within reserves, F , to preserve the species, and adjust the fraction of coastline in reserves, c , to maximize the number of larvae that settle outside the reserves (available as yield)’. Note that because F remains constant (something the modelers have chosen to assume), changing c means changing the width of reserves. Suppose that a value of F of 0.35 is deemed necessary to maintain the species. The solid line in Figure 15.20b shows how c and reserve width need to change to maintain an F of 0.35. The mathematical details of the model need not concern us but it turns out that although the largest yield is obtained when the reserves are as small as possible (the arrow in Figure 15.20b), so that larval export to fished areas is maximized, the yield is only slightly reduced as the reserve configuration moves away from this optimum. Thus, Hastings and Botsford (2003) argue that practical considerations, such as making reserves large enough to be enforced, can be allowed to play a major part in reserve design, as long as reserves are not so large (beyond the ‘shoulder’ of the curve in Figure 15.20b) as to significantly depress yield.

Although the model is a gross simplification, particularly in terms of the lack of any uncertainty or temporal or spatial heterogeneity, it usefully highlights some general considerations of importance and provides a starting point for more sophisticated and species-specific models to address the question of whether reserve networks will be useful for fisheries management.

In each of the sections of this chapter we have sought to build on relatively simple concepts by gradually adding more elements of realism. However, it should be remembered that even our most complex examples still lack realism in terms of the web of species interactions within which the target species are embedded. In fact, many management solutions have to be focused at a higher level of ecological organization – multispecies communities and whole ecosystems. We deal with the ecology of communities and ecosystems in Chapters 16–21 before considering ecological applications at this ecological level in Chapter 22.

Summary

Sustainability is a core concept in an ever-broadening concern for the fate of the earth and the ecological communities that occupy it. In this chapter we deal with two key aspects of ecological management – the control of pests and the management of harvests of wild populations. Each depends on an understanding of population interactions (discussed in Chapters 8–14) and each has sustainability as a primary aim.

One might imagine that the goal of pest control is total eradication but this is generally restricted to cases where a new pest has invaded a region and a rapid effort is made to completely eliminate it. Usually, the aim is to reduce the pest population to a level at which it does not pay to achieve yet more control (the economic injury level or EIL). In this way, we can see that economics and sustainability are intimately tied together. When a pest population has reached a density at which it is causing economic injury, however, it is generally too late to start controlling it. More important, then, is the economic threshold (ET): the density of the pest at which action should be taken to prevent it reaching the EIL.

We describe the tool kit of chemical pesticides and herbicides. These are a key part of the armory of pest managers but they

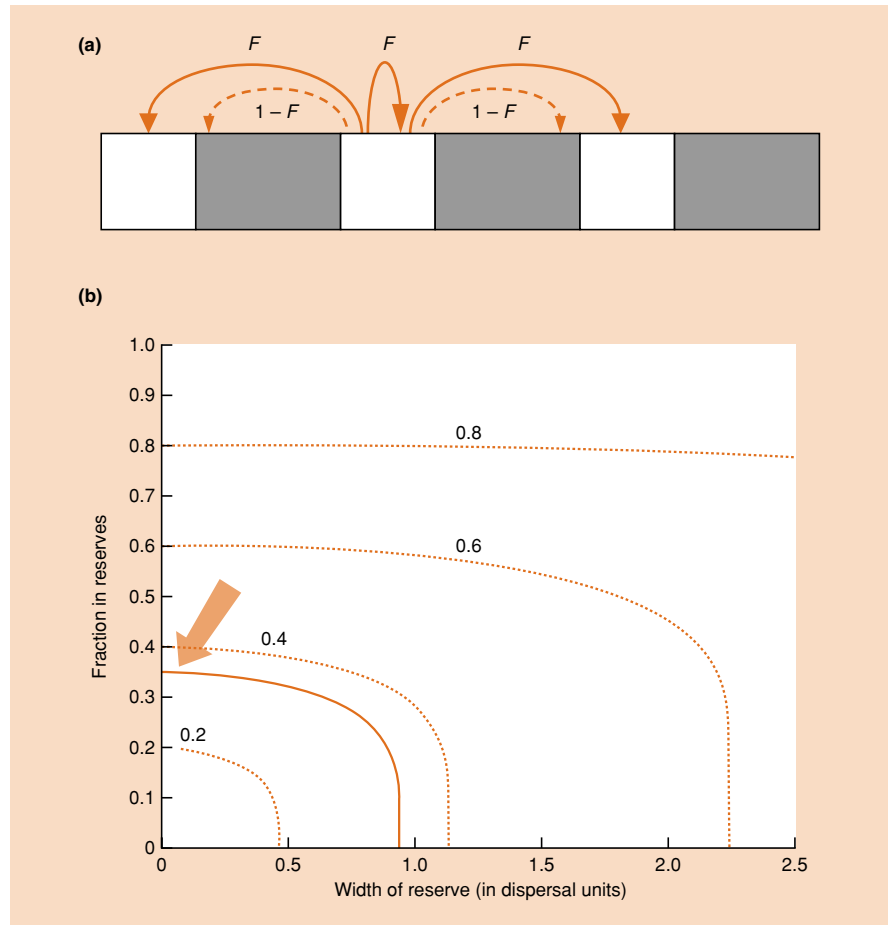


Figure 15.20 (a) Schematic representation of a network of marine reserves (white) and fished areas (gray). The fraction of coastline in the reserves is c , the fraction of larvae produced in the reserves is F , and the fraction of larvae produced in reserves that are exported is $1 - F$. (b) The combination of values of the fraction of coastline in reserves, c , and mean width of reserve (in units of mean dispersal distance) that yield a value of 0.35 for the fraction of larvae that are retained within reserves, F , along with similar combinations for other values of F . The arrow indicates the configuration that produces the maximum fishing yield outside the reserves. (After Hastings & Botsford, 2003.)

have to be used with care because of the possibility of ‘target pest resurgence’ (when treatment affects the pest’s natural enemies more than the pest itself) and ‘secondary pest outbreaks’ (when natural enemies of ‘potential’ pests are strongly affected, allowing potential pests to become actual pests). Pests are also adept at evolving resistance to pesticides.

An alternative to chemical pesticides is biological manipulation of the natural enemies of the pests. Biological control may involve: (i) ‘introduction’, with the expectation of long-term persistence, of a natural enemy from another geographic area (often the one from where the pest originated); (ii) the manipulation of natural predators already present (‘conservation biological control’); (iii) the periodic release of an agent that is unable to persist through the year but provides control for one or a few pest generations (‘inoculation’); or (iv) the release of large numbers of enemies, which will not persist, to kill only those pests present at the time (‘inundation’, sometimes called, by analogy, biological pesticides). Biological control is by no means always environmentally friendly. Examples are coming to light where even carefully chosen and apparently successful introductions of

biological control agents have impacted on nontarget species, both by affecting nontarget species related to the pest and by affecting other species that interact in food webs with the nontarget species.

Integrated pest management (IPM) is a practical philosophy of pest management that is ecologically based but uses all methods of control, including chemicals, when appropriate. It relies heavily on natural mortality factors such as weather and natural enemies.

Whenever, a natural population is exploited by harvesting there is a risk of overexploitation. But harvesters also want to avoid underexploitation, where potential consumers are deprived and those who harvest are underemployed. Thus, as with many areas of applied ecology, there are important economic and sociopolitical perspectives to consider.

The concept of the maximum sustainable yield (MSY) has been a guiding principle in harvest management. We describe the different approaches to obtain an MSY – taking a fixed quota, regulating harvest effort, harvesting a constant proportion or allowing constant escapement – and we point out the

shortcomings of each. More reliable approaches to sustainable harvesting are also discussed, including dynamic pool models (which recognize that all individuals in the harvested population are not equivalent and incorporate population structure into the population models) and approaches that explicitly incorporate economic factors (dealing with economically optimum yield, OEY, rather than simply MSY). We also note that no data are available for many of the world's fisheries, especially in developing areas

of the world; in these cases, simple 'dataless' management principles may be the best that ecologists can propose.

Finally, many populations, including those of pests and harvested populations, exist in a heterogeneous environment, sometimes as metapopulations. Managers need to be aware of this possibility, for instance when determining which biological control agent to use in an agricultural landscape or when designing a network of 'no-take' zones as part of a fisheries management strategy.