5.1 Introduction

Organisms grow, reproduce and die (Chapter 4). They are affected by the conditions in which they live (Chapter 2), and by the resources that they obtain (Chapter 3). But no organism lives in isolation. Each, for at least part of its life, is a member of a population composed of individuals of its own species.

Individuals of the same species have very similar requirements for survival, growth and reproduction; but their combined demand for a resource may exceed the immediate supply. The individuals then compete for the resource and, not surprisingly, at least some of them become deprived. This chapter is concerned with the nature of such intraspecific competition, its effects on the competing individuals and on populations of competing individuals. We begin with a working definition: ‘competition is an interaction between individuals, brought about by a shared requirement for a resource, and leading to a reduction in the survivorship, growth and/or reproduction of at least some of the competing individuals concerned’. We can now look more closely at competition.

Consider, initially, a simple hypothetical community: a thriving population of grasshoppers (all of one species) feeding on a field of grass (also of one species). To provide themselves with energy and material for growth and reproduction, grasshoppers eat grass; but in order to find and consume that grass they must use energy. Any grasshopper might find itself at a spot where there is no grass because some other grasshopper has eaten it. The grasshopper must then move on and expend more energy before it takes in food. The more grasshoppers there are, the more often this will happen. An increased energy expenditure and a decreased rate of food intake may all decrease a grasshopper’s chances of survival, and also leave less energy available for development and reproduction. Survival and reproduction determine a grasshopper’s contribution to the next generation. Hence, the more intraspecific competitors for food a grasshopper has, the less its likely contribution will be.

As far as the grass itself is concerned, an isolated seedling in fertile soil may have a very high chance of surviving to reproductive maturity. It will probably exhibit an extensive amount of modular growth, and will probably therefore eventually produce a large number of seeds. However, a seedling that is closely surrounded by neighbors (shading it with their leaves and depleting the water and nutrients of its soil with their roots) will be very unlikely to survive, and if it does, will almost certainly form few modules and set few seeds.

We can see immediately that the ultimate effect of competition on an individual is a decreased contribution to the next generation compared with what would have happened had there been no competitors. Intraspecific competition typically leads to decreased rates of resource intake per individual, and thus to decreased rates of individual growth or development, or perhaps to decreases in the amounts of stored reserves or to increased risks of predation. These may lead, in turn, to decreases in survivorship and/or decreases in fecundity, which together determine an individual's reproductive output.

5.1.1 Exploitation and interference

In many cases, competing individuals do not interact with one another directly. Instead, individuals respond to the level of a resource, which has been depressed by the presence and activity of other individuals. The grasshoppers were one example. Similarly, a competing grass plant is adversely affected by the presence of close neighbors, because the zone from which it extracts resources (light, water, nutrients) has been overlapped by the ‘resource depletion zones’ of these neighbors, making it more difficult to extract those resources. In such cases, competition may be described as...
Intraspecific competition, in that each individual is affected by the amount of resource that remains after that resource has been exploited by others. Exploitation can only occur, therefore, if the resource in question is in limited supply.

In many other cases, competition takes the form of interference. Here individuals interact directly with each other, and one individual will actually prevent another from exploiting the resources within a portion of the habitat. For instance, this is seen amongst animals that defend territories (see Section 5.11) and amongst the sessile animals and plants that live on rocky shores. The presence of a barnacle on a rock prevents any other barnacle from occupying that same position, even though the supply of food at that position may exceed the requirements of several barnacles. In such cases, space can be seen as a resource in limited supply. Another type of interference competition occurs when, for instance, two red deer stags fight for access to a harem of hinds. Either stag, alone, could readily mate with all the hinds, but they cannot both do so since matings are limited to the ‘owner’ of the harem.

Thus, interference competition may occur for a resource of real value (e.g. space on a rocky shore for a barnacle), in which case the interference is accompanied by a degree of exploitation, or for a surrogate resource (a territory, or ownership of a harem), which is only valuable because of the access it provides to a real resource (food, or females). With exploitation, the intensity of competition is closely linked to the level of resource present and the level required, but with interference, intensity may be high even when the level of the real resource is not limiting.

In practice, many examples of competition probably include elements of both exploitation and interference. For instance, adult cave beetles, Neapheanops tellkampfi, in Great Onyx Cave, Kentucky, compete amongst themselves but with no other species and have only one type of food – cricket eggs, which they obtain by digging holes in the sandy floor of the cave. On the one hand, they suffer indirectly from exploitation: beetles reduce the density of cricket eggs and then have markedly lower fecundity when food availability is low (Figure 5.1a). But they also suffer directly from interference: at higher beetle densities they fight more, forage less, dig fewer and shallower holes and eat far fewer eggs than could be accounted for by food depletion alone (Figure 5.1b).

### 5.1.2 One-sided competition

Whether they compete through exploitation or interference, individuals within a species have many fundamental features in common, using similar resources and reacting in much the same way to conditions. None the less, intraspecific competition may be very one sided: a strong, early seedling will shade a stunted, late one; an older and larger bryozoan on the shore will grow over a smaller and younger one. One example is shown in Figure 5.2. The overwinter survival of red deer calves in the resource-limited population on the island of Rhum, Scotland (see Chapter 4) declined sharply as the population became more crowded, but those that were smallest at birth were by far the most likely to die. Hence, the ultimate effect of competition is
far from being the same for every individual. Weak competitors may make only a small contribution to the next generation, or no contribution at all. Strong competitors may have their contribution only negligibly affected.

Finally, note that the likely effect of intraspecific competition on any individual is greater the more competitors there are. The effects of intraspecific competition are thus said to be density dependent. We turn next to a more detailed look at the density-dependent effects of intraspecific competition on death, birth and growth.

5.2 Intraspecific competition, and density-dependent mortality and fecundity

Figure 5.3 shows the pattern of mortality in the flour beetle Tribolium confusum when cohorts were reared at a range of densities. Known numbers of eggs were placed in glass tubes with 0.5 g of a flour–yeast mixture, and the number of individuals that survived to become adults in each tube was noted. The same data have been expressed in three ways, and in each case the resultant curve has been divided into three regions. Figure 5.3a describes the relationship between density and the per capita mortality rate – literally, the mortality rate ‘per head’, i.e. the probability of an individual dying or the proportion that died between the egg and adult stages. Figure 5.3b describes how the number that died prior to the adult stage changed with density; and Figure 5.3c describes the relationship between density and the numbers that survived.

Throughout region 1 (low density) the mortality rate remained constant as density was increased (Figure 5.3a). The numbers dying and the numbers surviving both rose (Figure 5.3b, c) (not surprising, given that the numbers ‘available’ to die and survive increased), but the proportion dying remained the same, which accounts for the straight lines in region 1 of these figures. Mortality in this region is said to be density independent. Individuals died, but the chance of an individual surviving to become an adult was not changed by the initial density. Judged by this, there was no intraspecific competition between the beetles at these densities. Such density-independent deaths affect the population at all densities. They represent a baseline, which any density-dependent mortality will exceed.

In region 2, the mortality rate increased with density (Figure 5.3a): there was density-dependent mortality. The numbers dying continued to rise with density, but unlike region 1 they did so more than proportionately (Figure 5.3b). The numbers surviving also continued to rise, but this time less than proportionately (Figure 5.3c). Thus, over this range, increases in egg density continued to lead to increases in the total number of surviving adults. The mortality rate had increased, but it ‘undercompensated’ for increases in density.

In region 3, intraspecific competition was even more intense. The increasing mortality rate ‘overcompensated’ for any increase in density, i.e. over this range, the more eggs there were present, the fewer adults survived: an increase in the initial number of eggs led to an even
greater proportional increase in the mortality rate. Indeed, if the range of densities had been extended, there would have been tubes with no survivors: the developing beetles would have eaten all the available food before any of them reached the adult stage.

A slightly different situation is shown in Figure 5.4. This illustrates the relationship between density and mortality in young trout. At the lower densities there was undercompensating density dependence, but at higher densities mortality never overcompensated. Rather, it compensated exactly for any increase in density: any rise in the number of fry was matched by an exactly equivalent rise in the mortality rate. The number of survivors therefore approached and maintained a constant level, irrespective of initial density.

The patterns of density-dependent fecundity that result from intraspecific competition are, in a sense, a mirror image of those for mortality (Figure 5.5). Here, though, the per capita birth rate falls as intraspecific competition intensifies. At low enough densities, the birth rate may be density independent (Figure 5.5a, lower densities). But as density increases, and the effects of intraspecific competition become apparent, birth rate initially shows undercompensating density dependence (Figure 5.5a, higher densities), and may then show exactly compensating density dependence (Figure 5.5b, throughout; Figure 5.5c, lower densities) or overcompensating density dependence (Figure 5.5c, higher densities).

Thus, to summarize, irrespective of variations in over- and undercompensation, the essential point is a simple one: at appropriate densities, intraspecific competition can lead to density-dependent mortality and/or fecundity, which means that the death rate increases and/or the birth rate decreases as density increases. Thus, whenever there is intraspecific competition, its effect, whether on survival, fecundity or a combination of the two, is density dependent. However, as subsequent chapters will show, there are processes other than intraspecific competition that also have density-dependent effects.

### 5.3 Density or crowding?

Of course, the intensity of intraspecific competition experienced by an individual is not really determined by the density of the population as a whole. The effect on an individual is determined,
rather, by the extent to which it is crowded or inhibited by its immediate neighbors.

One way of emphasizing this is by noting that there are actually at least three different meanings of ‘density’ (see Lewontin & Levins, 1989, where details of calculations and terms can be found). Consider a population of insects, distributed over a population of plants on which they feed. This is a typical example of a very general phenomenon – a population (the insects in this case) being distributed amongst different patches of a resource (the plants). The density would usually be calculated as the number of insects (let us say 1000) divided by the number of plants (say 100), i.e. 10 insects per plant. This, which we would normally call simply the ‘density’, is actually the ‘resource-weighted density’. However, it gives an accurate measure of the intensity of competition suffered by the insects (the extent to which they are crowded) only if there are exactly 10 insects on every plant and every plant is the same size.

Suppose, instead, that 10 of the plants support 91 insects each, and the remaining 90 support just one insect. The resource-weighted density would still be 10 insects per plant. But the average density experienced by the insects would be 82.9 insects per plant. That is, one adds up the densities experienced by each of the insects (91 + 91 + 91 . . . + 1 + 1) and divides by the total number of insects. This is the ‘organism-weighted density’, and it clearly gives a much more satisfactory measure of the intensity of competition the insects are likely to suffer.

However, there remains the further question of the average density of insects experienced by the plants. This, which may be referred to as the ‘exploitation pressure’, comes out at 1.1 insects per plant, reflecting the fact that most of the plants support only one insect.

What, then, is the density of the insect? Clearly, it depends on whether you answer from the perspective of the insect or the plant – but whichever way you look at it, the normal practice of calculating the resource-weighted density and calling it the ‘density’ looks highly suspect. The difference between resource- and organism-weighted densities is illustrated for the human population of a number of US states in Table 5.1 (where the ‘resource’ is simply land area). The organism-weighted densities are so much larger than the usual, but rather unhelpful, resource-weighted densities essentially because most people live, crowded, in cities (Lewontin & Levins, 1989).

The difficulties of relying on density to characterize the potential intensity of intraspecific competition are particularly
acute with sessile, modular organisms, because, being sessile, they compete almost entirely only with their immediate neighbors, and being modular, competition is directed most at the modules that are closest to those neighbors. Thus, for instance, when silver birch trees (Betula pendula) were grown in small groups, the sides of individual trees that interfaced with neighbors typically had a lower ‘birth’ and higher death rate of buds (see Section 4.2); whereas on sides of the same trees with no interference, bud birth rate was higher, death rate lower, branches were longer and the form approached that of an open-grown individual (Figure 5.6). Different modules experience different intensities of competition, and quoting the density at which an individual was growing would be all but pointless.

Thus, whether mobile or sessile, different individuals meet or suffer from different numbers of competitors. Density, especially resource-weighted density, is an abstraction that applies to the population as a whole but need not apply to any of the individuals within it. None the less, density may often be the most convenient way of expressing the degree to which individuals are crowded – and it is certainly the way it has usually been expressed.

<table>
<thead>
<tr>
<th>State</th>
<th>Resource-weighted density (km$^{-2}$)</th>
<th>Organism-weighted density (km$^{-2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorado</td>
<td>44</td>
<td>6,252</td>
</tr>
<tr>
<td>Missouri</td>
<td>159</td>
<td>6,525</td>
</tr>
<tr>
<td>New York</td>
<td>896</td>
<td>48,714</td>
</tr>
<tr>
<td>Utah</td>
<td>28</td>
<td>684</td>
</tr>
<tr>
<td>Virginia</td>
<td>207</td>
<td>13,824</td>
</tr>
</tbody>
</table>

Table 5.1  A comparison of the resource- and organism-weighted densities of five states, based on the 1960 USA census, where the ‘resource patches’ are the counties within each state. (After Lewontin & Levins, 1989.)

Figure 5.6  Mean relative bud production (new buds per existing bud) for silver birch trees (Betula pendula), expressed (a) as gross bud production and (b) as net bud production (birth minus death), in different interference zones. These zones are themselves explained in the inset.

- ●, high interference;
- ○, medium;
- ◆, low.

Bars represent standard errors. (After Jones & Harper, 1987.)
5.4 Intraspecific competition and the regulation of population size

There are, then, typical patterns in the effects of intraspecific competition on birth and death (see Figures 5.3–5.5). These generalized patterns are summarized in Figures 5.7 and 5.8.

5.4.1 Carrying capacities

Figure 5.7a–c reiterates the fact that as density increases, the per capita birth rate eventually falls and the per capita death rate eventually rises. There must, therefore, be a density at which these curves cross. At densities below this point, the birth rate exceeds

Figure 5.7 Density-dependent birth and mortality rates lead to the regulation of population size. When both are density dependent (a), or when either of them is (b, c), their two curves cross. The density at which they do so is called the carrying capacity \( K \). Below this the population increases, above it the population decreases: \( K \) is a stable equilibrium. However, these figures are the grossest of caricatures. The situation is closer to that shown in (d), where mortality rate broadly increases, and birth rate broadly decreases, with density. It is possible, therefore, for the two rates to balance not at just one density, but over a broad range of densities, and it is towards this broad range that other densities tend to move.

Figure 5.8 Some general aspects of intraspecific competition. (a) Density-dependent effects on the numbers dying and the number of births in a population: net recruitment is 'births minus deaths'. Hence, as shown in (b), the density-dependent effect of intraspecific competition on net recruitment is a domed or 'n'-shaped curve. (c) A population increasing in size under the influence of the relationships in (a) and (b). Each arrow represents the change in size of the population over one interval of time. Change (i.e. net recruitment) is small when density is low (i.e. at small population sizes: A to B, B to C) and is small close to the carrying capacity (I to J, J to K), but is large at intermediate densities (E to F). The result is an 'S'-shaped or sigmoidal pattern of population increase, approaching the carrying capacity.
the death rate and the population increases in size. At densities above the crossover point, the death rate exceeds the birth rate and the population declines. At the crossover density itself, the two rates are equal and there is no net change in population size. This density therefore represents a stable equilibrium, in that all other densities will tend to approach it. In other words, intraspecific competition, by acting on birth rates and death rates, can regulate populations at a stable density at which the birth rate equals the death rate. This density is known as the carrying capacity of the population and is usually denoted by \( K \) (Figure 5.7). It is called a carrying capacity because it represents the population size that the resources of the environment can just maintain (‘carry’) without a tendency to either increase or decrease.

However, whilst hypothetical populations caricatured by line drawings like Figures 5.7a–c can be characterized by a simple carrying capacity, this is not true of any natural population. There are unpredictable environmental fluctuations; individuals are affected by a whole wealth of factors of which intraspecific competition is only one; and resources not only affect density but respond to density as well. Hence, the situation is likely to be closer to that depicted in Figure 5.7d. Intraspecific competition does not hold natural populations to a predictable and unchanging level (the carrying capacity), but it may act upon a very wide range of starting densities and bring them to a much narrower range of final densities, and it therefore tends to keep density within certain limits. It is in this sense that intraspecific competition may be said typically to be capable of regulating population size. For instance, Figure 5.9 shows the fluctuations within and between years in populations of the brown trout (\textit{Salmo trutta}) and the grasshopper, \textit{Chorthippus brunneus}. There are no simple carrying capacities in these examples, but there are clear tendencies for the ‘final’ density each year (‘late summer numbers’ in the first case, ‘adults’ in the second) to be relatively constant, despite the large fluctuations in density within each year and the obvious potential for increase that both populations possess.

In fact, the concept of a population settling at a stable carrying capacity, even in caricatured populations, is relevant only to situations in which density dependence is not strongly overcompensating. Where there is overcompensation, cycles or even...
chaotic changes in population size may be the result. We return to this point later (see Section 5.8).

5.4.2 Net recruitment curves

An alternative general view of intraspecific competition is shown in Figure 5.8a, which deals with numbers rather than rates. The difference there between the two curves ('births minus deaths' or 'net recruitment') is the net number of additions expected in the population during the appropriate stage or over one interval of time. Because of the shapes of the birth and death curves, the net number of additions is small at the lowest densities, increases as density rises, declines again as the carrying capacity is approached and is then negative (deaths exceed births) when the initial density exceeds $K$ (Figure 5.8b). Thus, total recruitment into a population is small when there are few individuals available to give birth, and small when intraspecific competition is intense. It reaches a peak, i.e. the population increases in size most rapidly, at some intermediate density.

The precise nature of the relationship between a population's net rate of recruitment and its density varies with the detailed biology of the species concerned (e.g. the trout, clover plants, herring and whales in Figure 5.10a–d). Moreover, because recruitment is affected by a whole multiplicity of factors, the data points rarely fall exactly on any single curve. Yet, in each case in Figure 5.10, a domed curve is apparent. This reflects the general nature of density-dependent birth and death whenever there is intraspecific competition. Note also that one of these (Figure 5.10b) is modular: it describes the relationship between the leaf area index (LAI) of a plant population (the total leaf area being borne per unit area of ground) and the population's growth rate (modular birth minus modular death). The growth rate is low when there are few leaves, peaks at an intermediate LAI, and is then low again at a high LAI, where there is much mutual shading and competition and many leaves may be consuming more in respiration than they contribute through photosynthesis.

5.4.3 Sigmoidal growth curves

In addition, curves of the type shown in Figure 5.8a and b may be used to suggest the pattern by which a population might increase from an initially very small size (e.g. when a species colonizes a previously unoccupied area). This is illustrated in Figure 5.8c. Imagine a small population, well below the carrying capacity of its environment (point A). Because the population is small, it increases in size only slightly during one time interval, and only reaches point B. Now, however, being larger, it increases in size more rapidly during the next time interval (to point C), and even more during the next (to point D). This process continues until the population passes beyond the peak of its net recruitment curve (Figure 5.8b). Thereafter, the population increases in size less and less with each time interval until the population reaches its

![Figure 5.10](https://example.com/5.10.png)

intrinsically carrying capacity \( (K) \) and ceases completely to increase in size. The population might therefore be expected to follow an S-shaped or 'sigmoidal' curve as it rises from a low density to its carrying capacity. This is a consequence of the hump in its recruitment rate curve, which is itself a consequence of intraspecific competition.

Of course, Figure 5.8c, like the rest of Figure 5.8, is a gross simplification. It assumes, apart from anything else, that changes in population size are affected only by intraspecific competition. Nevertheless, something akin to sigmoidal population growth can be perceived in many natural and experimental situations (Figure 5.11).

Intraspecific competition will be obvious in certain cases (such as overgrowth competition between sessile organisms on a rocky shore), but this will not be true of every population examined. Individuals are also affected by predators, parasites and prey, competitors from other species, and the many facets of their physical and chemical environment. Any of these may outweigh or obscure the effects of intraspecific competition; or the effect of these other factors at one stage may reduce the density to well below the carrying capacity for all subsequent stages. Nevertheless, intraspecific competition probably affects most populations at least sometimes during at least one stage of their life cycle.

### 5.5 Intraspecific competition and density-dependent growth

Intraspecific competition, then, can have a profound effect on the number of individuals in a population; but it can have an equally profound effect on the individuals themselves. In populations of unitary organisms, rates of growth and rates of development are commonly influenced by intraspecific competition. This necessarily leads to density-dependent effects on the composition of a population. For instance, Figure 5.12a and b shows two examples in which individuals were typically smaller at higher densities. This, in turn, often means that although the numerical size of a population is regulated only approximately by intraspecific competition, the total biomass is regulated much more precisely. This, too, is illustrated by the limpets in Figure 5.12b.

#### 5.5.1 The law of constant final yield

Such effects are particularly marked in modular organisms. For example, when carrot seeds (Daucus carota) were sown at a range of densities, the yield per pot at the first harvest (29 days) increased with the density of seeds sown (Figure 5.13). After 62 days, however, and even more after 76 and 90 days, yield no longer reflected the numbers sown. Rather it was the same over a wide range of initial densities, especially at higher densities where competition was most intense. This pattern has frequently been noted by plant ecologists and has been called the 'law of constant final yield' (Kira et al., 1953). Individuals suffer density-dependent reductions in growth rate, and thus in individual plant size, which tend to compensate exactly for increases in density (hence the constant final yield). This suggests, of course, that there are limited resources available for plant growth, especially at high densities, which is borne out in Figure 5.13 by the higher (constant) yields at higher nutrient levels.
Yield is density \( d \) multiplied by mean weight per plant \( \bar{w} \).

Thus, if yield is constant \( c \):

\[ d\bar{w} = c \]

and so:

\[ \log d + \log \bar{w} = \log c \]  \hspace{1cm} (5.1)

and:

\[ \log \bar{w} = \log c - 1 \cdot \log d \]  \hspace{1cm} (5.2)

and thus, a plot of log mean weight against log density should have a slope of \(-1\).

Data on the effects of density on the growth of the grass *Vulpia fasciculata* are shown in Figure 5.14, and the slope of the curve towards the end of the experiment does indeed approach a value of \(-1\). Here too, as with the carrot plants, individual plant weight at the first harvest was reduced only at very high densities – but as the plants became larger, they interfered with each other at successively lower densities.

The constancy of the final yield is a result, to a large extent, of the modularity of plants. This was clear when perennial rye grass (*Lolium perenne*) was sown at a 30-fold range of densities (Figure 5.15). After 180 days some genets had died; but the range of final tiller (module) densities was far narrower than that of genets (individuals). The regulatory powers of intraspecific competition were operating largely by affecting the number of modules per genet rather than the number of genets themselves.

### 5.6 Quantifying intraspecific competition

Every population is unique. Nevertheless, we have already seen that there are general patterns in the action of intraspecific competition. In this section we take such generalizations a stage further. A method will be described, utilizing \( k \) values (see Chapter 4) to summarize the effects of intraspecific competition on mortality, fecundity and growth. Mortality will be dealt with first. The method will then be extended for use with fecundity and growth.

A \( k \) value was defined by the formula:

\[ k = \log (\text{initial density}) - \log (\text{final density}), \]  \hspace{1cm} (5.4)

or, equivalently:

\[ k = \log (\text{initial density}/\text{final density}). \]  \hspace{1cm} (5.5)

For present purposes, ‘initial density’ may be denoted by \( B \), standing for ‘numbers before the action of intraspecific competition’, whilst ‘final density’ may be denoted by \( A \), standing for ‘numbers after the action of intraspecific competition’. Thus:

\[ k = \log (B/A). \]  \hspace{1cm} (5.6)

Note that \( k \) increases as mortality rate increases.

Some examples of the effects of intraspecific competition on mortality are shown in Figure 5.16, in which \( k \) is plotted against log \( B \). In several cases, \( k \) is constant at the lowest densities. This is an indication of density independence: the proportion surviving is not correlated with initial density. At higher densities, \( k \) increases with initial density; this indicates density dependence. Most importantly,
however, the way in which $k$ varies with the logarithm of density indicates the precise nature of the density dependence. For example, Figure 5.16a and b describes, respectively, situations in which there is under- and exact compensation at higher densities. The exact compensation in Figure 5.16b is indicated by the slope of the curve (denoted by $b$) taking a constant value of 1 (the mathematically inclined will see that this follows from the fact that with exact compensation, $A$ is constant). The undercompensation that preceded this at lower densities, and which is seen in Figure 5.16a even at higher densities, is indicated by the fact that $b$ is less than 1.

Exact compensation ($b = 1$) is often referred to as pure contest competition, because there are a constant number of winners (survivors) in the competitive process. The term was initially proposed by Nicholson (1954), who contrasted it with what he called pure scramble competition. Pure scramble is the most extreme form of overcompensating density dependence, in which all competing individuals are so adversely affected that none of them survive, i.e. $A = 0$. This would be indicated in Figure 5.16 by a $b$ value of infinity (a vertical line), and Figure 5.16c is an example in which this is the case. More common, however, are examples in which competition is scramble-like, i.e. there is considerable but not total overcompensation ($b > 1$). This is shown, for instance, in Figure 5.16d.

Plotting $k$ against log $B$ is thus an informative way of depicting the effects of intraspecific competition on mortality. Variations in the slope of the curve ($b$) give a clear indication

---

**Figure 5.13** The relationship between yield per pot and sowing density in carrots (*Daucus carota*) at four harvests ((a) 29 days after sowing, (b) 62 days, (c) 76 days, and (d) 90 days) and at three nutrient levels (low, medium and high: L, M and H), given to pots weekly after the first harvest. Points are means of three replicates, with the exception of the lowest density (9) and the first harvest (9). ○, root weight; ●, shoot weight; □, total weight. The curves were fitted in line with theoretical yield–density relationships, the details of which are unimportant in this context. (After Li et al., 1996.)
Figure 5.14  The ‘constant final yield’ of plants illustrated by a line of slope $-1$ when log mean weight is plotted against log density in the dune annual, *Vulpia fasciculata*. On January 18, particularly at low densities, growth and hence mean dry weight were roughly independent of density. But by June 27, density-dependent reductions in growth compensated exactly for variations in density, leading to a constant yield. (After Watkinson, 1984.)

Figure 5.15  Intraspecific competition in plants often regulates the number of modules. When populations of rye grass (*Lolium perenne*) were sown at a range of densities, the range of final tiller (i.e. module) densities was far narrower than that of genets. (After Kays & Harper, 1974.)

of the manner in which density dependence changes with density. The method can also be extended to fecundity and growth.

For fecundity, it is necessary to think of $B$ as the ‘total number of offspring that would have been produced had there been no intraspecific competition’, i.e. if each reproducing individual had produced as many offspring as it would have done in a competition-free environment. $A$ is then the total number of offspring actually produced. (In practice, $B$ is usually estimated from the population experiencing the least competition – not necessarily competition-free.) For growth, $B$ must be thought of as the total biomass, or total number of modules, that would have been produced had all individuals grown as if they were in a competition-free situation. $A$ is then the total biomass or total number of modules actually produced.

Figure 5.17 provides examples in which $k$ values are used to describe the effects of intraspecific competition on fecundity and growth. The patterns are essentially similar to those in Figure 5.16. Each falls somewhere on the continuum ranging between density independence and pure scramble, and their position along that continuum is immediately apparent. Using $k$ values, all examples of intraspecific competition can be quantified in the same terms. With fecundity and growth, however, the terms ‘scramble’ and especially ‘contest’ are less appropriate. It is better simply to talk in terms of exact, over- and undercompensation.

### 5.7 Mathematical models: introduction

The desire to formulate general rules in ecology often finds its expression in the construction of mathematical or graphical models. It may seem surprising that those interested in the natural living world should spend time reconstructing it in an artificial mathematical form; but there are several good reasons why this should be done. The first is that models can crystallize, or at least bring together in terms of a few parameters, the important, shared properties of a wealth of unique examples. This simply makes it easier for ecologists to think about the problem or process under consideration, by forcing us to try to extract the essentials from complex systems. Thus, a model can provide a ‘common language’ in which each unique example can be expressed; and if each can be expressed in a common language, then their properties relative to one another, and relative perhaps to some ideal standard, will be more apparent.

These ideas are more familiar, perhaps, in other contexts. Newton never laid hands on a perfectly frictionless body, and Boyle never saw an ideal gas – other than in their imaginations – but Newton’s Laws of Motion and Boyle’s Law have been of immeasurable value to us for centuries.

Perhaps more importantly, however, models can actually shed light on the real world that they mimic. Specific examples below will make this apparent. Models can, as we shall see, exhibit properties that the system being modeled had not previously been known to possess. More commonly, models make it clear how the behavior of a population, for example, depends on the properties of the individuals that comprise it. That is, models allow us to see the likely consequences of any assumptions that we choose to make – ‘If it were the case that only juveniles migrate, what would this do to the dynamics of their populations?’ – and so on. Models can do this because mathematical methods are designed precisely to allow a set of assumptions to be followed through
to their natural conclusions. As a consequence, models often suggest what would be the most profitable experiments to carry out or observations to make – ‘Since juvenile migration rates appear to be so important, these should be measured in each of our study populations’.

These reasons for constructing models are also criteria by which any model should be judged. Indeed, a model is only useful (i.e. worth constructing) if it does perform one or more of these functions. Of course, in order to perform them a model must adequately describe real situations and real sets of data, and this ‘ability to describe’ or ‘ability to mimic’ is itself a further criterion by which a model can be judged. However, the crucial word is ‘adequate’. The only perfect description of the real world is the real world itself. A model is an adequate description, ultimately, as long as it performs a useful function.

In the present case, some simple models of intraspecific competition will be described. They will be built up from a very elementary starting point, and their properties (i.e. their ability to satisfy the criteria described above) will then be examined. Initially, a model will be constructed for a population with discrete breeding seasons.

### 5.8 A model with discrete breeding seasons

#### 5.8.1 Basic equations

In Section 4.7 we developed a simple model for species with discrete breeding seasons, in which the population size at time \( t \), \( N_t \), altered in size under the influence of a fundamental net reproductive rate, \( R \). This model can be summarized in two equations:

\[
N_{t+1} = N_t R
\]  
(5.7)

and:

\[
N_t = N_0 R^t.
\]  
(5.8)

The model, however, describes a population in which there is no competition. \( R \) is constant, and if \( R > 1 \), the population will continue to increase in size indefinitely (‘exponential growth’, shown in Figure 5.18). The first step is therefore to modify the equations by making the net reproductive rate subject to intraspecific competition. This is done in Figure 5.19, which has three components.

At point A, the population size is very small (\( N_t \) is virtually zero). Competition is therefore negligible, and the actual net reproductive rate is adequately defined by an unmodified \( R \). Thus, Equation 5.7 is still appropriate, or, rearranging the equation:

\[
N_t/N_{t+1} = 1/R.
\]  
(5.9)

At point B, by contrast, the population size (\( N_t \)) is very much larger and there is a significant amount of intraspecific competition, such that the net reproductive rate has been so modified by competition that the population can collectively do no better than replace itself each generation, because ‘births’ equal ‘deaths’. In other words, \( N_{t+1} \) is simply the same as \( N_t \), and \( N_t/N_{t+1} \) equals 1. The population size at which this occurs is, by definition, the carrying capacity, \( K \) (see Figure 5.7).

The third component of Figure 5.19 is the straight line joining point A to point B and extending beyond it. This describes the progressive modification of the actual net reproductive rate as population size increases; but its straightness is simply an
assumption made for the sake of expediency, since all straight lines are of the simple form: \( y = (\text{slope}) \times x + (\text{intercept}) \). In Figure 5.19, \( N_t/N_{r1} \) is measured on the y-axis, \( N_t \) on the x-axis, the intercept is \( 1/R \) and the slope, based on the segment between points A and B, is \((1 - 1/R)/K\). Thus:

\[
\frac{N_t}{N_{r1}} = \frac{1}{K} \cdot \frac{1}{N_t} + \frac{1}{R}
\]

or, rearranging:

\[
N_{r1} = \frac{N_tR}{1 + (R-1)N_t/K}
\]

For further simplicity, \((R-1)/K\) may be denoted by \(a\) giving:

\[
N_{r1} = \frac{N_tR}{1 + aN_t}
\]

This is a model of population increase limited by intraspecific competition. Its essence lies in the fact that the unrealistically constant \(R\) in Equation 5.7 has been replaced by an actual net reproductive rate, \(R/(1 + aN_t)\), which decreases as population size \(N_t\) increases.

We, like many others, derived Equation 5.12 as if the behavior of a population is jointly determined by \(R\) and \(K\), the per capita rate of increase and the population’s carrying capacity – \(a\) is then simply a particular combination of these. An alternative point of view is that \(a\) is meaningful in its own right, measuring the per capita susceptibility to crowding: the larger the value of \(a\), the greater the effect of density on the actual rate of increase in the population (Kuno, 1991). Now the behavior of a population is seen as being jointly determined by two properties of the individuals within it – their intrinsic per capita rate of increase and their susceptibility to crowding, \(R\) and \(a\). The carrying capacity of the population \((K = (R-1)/a)\) is then simply an outcome of these properties. The great advantage of this viewpoint is that it places individuals and populations in a more realistic biological perspective. Individuals come first: individual birth rates, death rates and susceptibilities to crowding are subject to natural selection and evolve. Populations simply follow: a population’s carrying capacity is just one of many features that reflect the values these individual properties take.

The properties of the model in Equation 5.12 may be seen in Figure 5.19 (from which the model was derived) and Figure 5.18 (which shows a hypothetical population increasing in size over time in conformity with the model). The population in Figure 5.18 describes an S-shaped curve over time. As we saw earlier, this is a desirable quality of a model of intraspecific competition. Note, however, that there are many other models that would also generate such a curve. The advantage of Equation 5.12 is its simplicity.

The behavior of the model in the vicinity of the carrying capacity can best be seen by reference to Figure 5.19. At population sizes that are less than \(K\) the population will increase in size; at population sizes that are greater than \(K\) the population size will decline; and at \(K\) itself the population neither increases nor decreases. The carrying capacity is therefore a stable equilibrium for the population, and the model exhibits the regulatory properties classically characteristic of intraspecific competition.

### 5.8.2 What type of competition?

It is not yet clear, however, just exactly what type or range of competition this model is able to describe. This can be explored by tracing the relationship between \(k\) values and \(\log N\) (as in Section 5.6). Each generation, the potential number of individuals produced (i.e. the number that would be produced if there were no competition) is \(N_tR\). The actual number produced (i.e. the number that survive the effects of competition) is \(N_tR/(1 + aN_t)\).

Section 5.6 established that:

\[
k = \log (\text{number produced}) - \log (\text{number surviving})
\]

Thus, in the present case:

\[
k = \log N_tR - \log N_tR/(1 + aN_t),
\]

or, simplifying:

\[
k = \log(1 + aN_t).
\]

Figure 5.20 shows a number of plots of \(k\) against \(\log N\), with a variety of values of \(a\) inserted into the model. In every case, the slope of the graph approaches and then attains a value of 1. In other words, the density dependence always begins by undercompensating and then compensates perfectly at higher values of \(N\). The model is therefore limited in the type of competition that it can produce, and all we have been able to say so far is that this type of competition leads to very tightly controlled regulation of populations.

### 5.8.3 Time lags

One simple modification that we can make is to relax the assumption that populations respond instantaneously to changes
in their own density, i.e. that present density determines the amount of resource available to a population and this in turn determines the net reproductive rate within the population. Suppose instead that the amount of resource available is determined by the density one time interval previously. To take a specific example, the amount of grass in a field in spring (the resource available to cattle) might be determined by the level of grazing (and hence, the density of cattle) in the previous year. In such a case, the reproductive rate itself will be dependent on the density one time interval ago. Thus, since in Equations 5.7 and 5.12:

\[ N_{t+1} = N_t \times \text{reproductive rate}, \]

Equation 5.12 may be modified to:

\[ N_{t+1} = \frac{N_t R}{1 + aN_t b}. \]  \hspace{1cm} (5.17)

There is a time lag in the population’s response to its own density, caused by a time lag in the response of its resources. The behavior of the modified model is as follows:

\[ R < 1.33: \text{direct approach to a stable equilibrium} \]
\[ R > 1.33: \text{damped oscillations towards that equilibrium}. \]

In comparison, the original Equation 5.12, without a time lag, gave rise to a direct approach to its equilibrium for all values of \( R \). The time lag has provoked the fluctuations in the model, and it can be assumed to have similar, destabilizing effects on real populations.

5.8.4 Incorporating a range of competition

A simple modification of Equation 5.12 of far more general importance was originally suggested by Maynard Smith and Slatkin (1973) and was discussed in detail by Bellows (1981). It alters the equation to:

\[ N_{t+1} = \frac{N_t R}{1 + aN_t b}. \]  \hspace{1cm} (5.18)

The justification for this modification may be seen by examining some of the properties of the revised model. For example, Figure 5.21 shows plots of \( k \) against \( \log_{10} N_t \), analogous to those in Figure 5.20: \( k \) is now \( \log_{10}[1 + (aN_t b)^b] \). The slope of the curve, instead of approaching 1 as it did previously, now approaches the value taken by \( b \) in Equation 5.18. Thus, by the choice of appropriate values, the model can portray undercompensation (\( b < 1 \)), perfect compensation (\( b = 1 \)), scramble-like overcompensation (\( b > 1 \)) or even density independence (\( b = 0 \)). This model has the generality that Equation 5.12 lacks, with the value of \( b \) determining the type of density dependence that is being incorporated.

Another desirable quality that Equation 5.18 shares with other good models is an ability to throw fresh light on the real world. By sensible analysis of the population dynamics generated by the equation,
it is possible to draw guarded conclusions about the dynamics of natural populations. The mathematical method by which this and similar equations may be examined is set out and discussed by May (1975a), but the results of the analysis (Figure 5.22) can be appreciated without dwelling on the analysis itself. Figure 5.22b shows the various patterns of population growth and dynamics that Equation 5.18 can generate. Figure 5.22a sets out the conditions under which each of these patterns occurs. Note first that the pattern of dynamics depends on two things: (i) \( b \), the precise type of competition or density dependence; and (ii) \( R \), the effective net reproductive rate (taking density-independent mortality into account). By contrast, \( a \) determines not the type of pattern, but only the level about which any fluctuations occur.

As Figure 5.22a shows, low values of \( b \) and/or \( R \) lead to populations that approach their equilibrium size without fluctuating at all (‘monotonic damping’). This has already been hinted at in Figure 5.18. There, a population behaving in conformity with Equation 5.12 approached equilibrium directly, irrespective of the value of \( R \). Equation 5.12 is a special case of Equation 5.18 in which \( b = 1 \) (perfect compensation); Figure 5.22a confirms that for \( b = 1 \), monotonic damping is the rule whatever the effective net reproductive rate.

As the values of \( b \) and/or \( R \) increase, the behavior of the population changes first to damped oscillations gradually approaching equilibrium, and then to ‘stable limit cycles’ in which the population fluctuates around an equilibrium level, revisiting the same two, four or even more points time and time again. Finally, with large values of \( b \) and \( R \), the population fluctuates in an apparently irregular and chaotic fashion.

### 5.8.5 Chaos

Thus, a model built around a density-dependent, supposedly regulatory process (intraspecific competition) can lead to a very wide range of population dynamics. If a model population has even a moderate fundamental net reproductive rate (and the ability to leave 100 (\( = R \)) offspring in the next generation in a competition-free environment is not unreasonable), and if it has a density-dependent reaction which even moderately overcompensates, then far from being stable, it may fluctuate wildly in numbers without the action of any extrinsic factor. The biological significance of this is the strong suggestion that even in an environment that is wholly constant and predictable, the intrinsic qualities of a population and the individuals within it may, by themselves, give rise to population dynamics with large and perhaps even chaotic fluctuations. The consequences of intraspecific competition are clearly not limited to ‘tightly controlled regulation’.

This leads us to two important conclusions. First, time lags, high reproductive rates and overcompensating density dependence are capable (either alone or in combination) of producing all types of fluctuations in population density, without invoking any extrinsic cause. Second, and equally important, this has been made apparent by the analysis of mathematical models.
In fact, the recognition that even simple ecological systems may contain the seeds of chaos has led to chaos itself becoming a topic of interest amongst ecologists (Schaffer & Kot, 1986; Hastings et al., 1993; Perry et al., 2000). A detailed exposition of the nature of chaos is not appropriate here, but a few key points should be understood.

1. The term ‘chaos’ may itself be misleading if it is taken to imply a fluctuation with absolutely no discernable pattern. Chaotic dynamics do not consist of a sequence of random numbers. On the contrary, there are tests (although they are not always easy to put into practice) designed to distinguish chaotic from random and other types of fluctuations.

2. Fluctuations in chaotic ecological systems occur between definable upper and lower densities. Thus, in the model of intraspecific competition that we have discussed, the idea of ‘regulation’ has not been lost altogether, even in the chaotic region.

3. Unlike the behavior of truly regulated systems, however, two similar population trajectories in a chaotic system will not tend to converge on (‘be attracted to’) the same equilibrium density or the same limit cycle (both of them ‘simple’ attractors). Rather, the behavior of a chaotic system is governed by a ‘strange attractor’. Initially, very similar trajectories will diverge from one another, exponentially, over time: chaotic systems exhibit ‘extreme sensitivity to initial conditions’.

4. Hence, the long-term future behavior of a chaotic system is effectively impossible to predict, and prediction becomes increasingly inaccurate as one moves further into the future. Even if we appear to have seen the system in a particular state before – and know precisely what happened subsequently last time – tiny (perhaps immeasurable) initial differences will be magnified progressively, and past experience will become of increasingly little value.

Ecology must aim to become a predictive science. Chaotic systems set us some of the sternest challenges in prediction. There has been an understandable interest, therefore, in the question ‘How often, if ever, are ecological systems chaotic?’ Attempts to answer this question, however, whilst illuminating, have certainly not been definitive.

Most recent attempts to detect chaos in ecological systems have been based on a mathematical advance known as Takens’ theorem. This says, in the context of ecology, that even when a system comprises a number of interacting elements, its characteristics (whether it is chaotic, etc.) may be deduced from a time series of abundances of just one of those elements (e.g. one species). This is called ‘reconstructing the attractor’. To be more specific: suppose, for example, that a system’s behavior is determined by interactions between four elements (for simplicity, four species). First, one expresses the abundance of just one of those species at time \( t \), \( N_t \), as a function of the sequence of abundances at \( four \) successive previous time points: \( N_{t-1}, N_{t-2}, N_{t-3}, N_{t-4} \) (the same number of ‘lags’ as there are elements in the original system). Then, the attractor of this lagged system of abundances is an accurate reconstruction of the attractor of the original system, which determines its characteristics.

In practice, this means taking a series of abundances of, say, one species and finding the ‘best’ model, in statistical terms, for predicting \( N_t \) as a function of lagged abundances, and then investigating this reconstructed attractor as a means of investigating the nature of the dynamics of the underlying system. Unfortunately, ecological time series (compared, say, to those of physics) are particularly short and particularly noisy. Thus, methods for identifying a ‘best’ model and applying Takens’ theorem, and for identifying chaos in ecology generally, have been ‘the focus of continuous methodological debate and refinement’ (Bjørnstad & Grenfell, 2001), one consequence of which is that any suggestion of a suitable method in a textbook such as this is almost certainly doomed to be outmoded by the time it is first read.

Notwithstanding these technical difficulties, however, and in spite of occasional demonstrations of apparent chaos in artificial laboratory environments (Costantino et al., 1997), a consensus view has grown that chaos is not a dominant pattern of dynamics in natural ecological systems. One trend, therefore, has been to seek to understand why chaos might not occur in nature, despite its being generated readily by ecological models. For example, Fussmann and Heber (2002) examined model populations embedded in food webs and found that as the webs took on more of the characteristics observed in nature (see Chapter 20) chaos became less likely.

Thus, the potential importance of chaos in ecological systems is clear. From a fundamental point of view, we need to appreciate that if we have a relatively simple system, it may nevertheless generate complex, chaotic dynamics; and that if we observe complex dynamics, the underlying explanation may nevertheless be simple. From an applied point of view, if ecology is to become a predictive and manipulative science, then we need to know the extent to which long-term prediction is threatened by one of the hallmarks of chaos – extreme sensitivity to initial conditions. The key practical question, however – ‘how common is chaos?’ – remains largely unanswered.

### 5.9 Continuous breeding: the logistic equation

The model derived and discussed in Section 5.8 was appropriate for populations that have discrete breeding seasons and can therefore be described by equations growing in discrete steps, i.e. by ‘difference’ equations. Such models are not appropriate,
however, for those populations in which birth and death are continuous. These are best described by models of continuous growth, or ‘differential’ equations, which will be considered next.

The net rate of increase of such a population will be denoted by \( \frac{dN}{dt} \) (referred to in speech as ‘dN by dr’). This represents the ‘speed’ at which a population increases in size, \( N \), as time, \( t \), progresses. The increase in size of the whole population is the sum of the contributions of the various individuals within it. Thus, the average rate of increase per individual, or the ‘per capita rate of increase’ is given by \( \frac{dN}{dt} \left(\frac{1}{N}\right) \). But we have already seen in Section 4.7 that in the absence of competition, this is the definition of the ‘intrinsic rate of natural increase’, \( r \). Thus:

\[
\frac{dN}{dt} \left(\frac{1}{N}\right) = r
\]  
(5.19)

and:

\[
\frac{dN}{dt} = rN.
\]  
(5.20)

A population increasing in size under the influence of Equation 5.20, with \( r > 0 \), is shown in Figure 5.23. Not surprisingly, there is unlimited, ‘exponential’ increase. In fact, Equation 5.20 is the continuous form of the exponential difference Equation 5.8, and as discussed in Section 4.7, \( r \) is simply \( \log R \). (Mathematically adept readers will see that Equation 5.20 can be obtained by differentiating Equation 5.8.) \( R \) and \( r \) are clearly measures of the same commodity: ‘birth plus survival’ or ‘birth minus death’; the difference between \( R \) and \( r \) is merely a change of currency.

For the sake of realism, intraspecific competition must obviously be added to Equation 5.20. This can be achieved most simply by a method exactly equivalent to the one used in Figure 5.19, giving rise to:

\[
\frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right).
\]  
(5.21)

This is known as the logistic equation (coined by Verhulst, 1838), and a population increasing in size under its influence is shown in Figure 5.23.

The logistic equation is the continuous equivalent of Equation 5.12, and it therefore has all the essential characteristics of Equation 5.12, and all of its shortcomings. It describes a sigmoidal growth curve approaching a stable carrying capacity, but it is only one of many reasonable equations that do this. Its major advantage is its simplicity. Moreover, whilst it was possible to incorporate a range of competitive intensities into Equation 5.12, this is by no means easy with the logistic equation. The logistic is therefore doomed to be a model of perfectly compensating density dependence. Nevertheless, in spite of these limitations, the equation will be an integral component of models in Chapters 8 and 10, and it has played a central role in the development of ecology.

5.10 Individual differences: asymmetric competition

5.10.1 Size inequalities

Until now, we have focused on what happens to the whole population or the average individual within it. Different individuals, however, may respond to intraspecific competition in very different ways. Figure 5.24 shows the results of an experiment in which flax (\( \text{Linum usitatissimum} \)) was sown at three densities, and harvested at three stages of development, recording the weight of each plant individually. This made it possible to monitor the effects of increasing amounts of competition not only as a result of variations in sowing density, but also as a result of plant growth (between the first and the last harvests). When intraspecific competition was at its least intense (at the lowest sowing density after only 2 weeks’ growth) the individual plant weights were distributed symmetrically about the mean. When competition was at its most intense, however, the distribution was strongly skewed to the left: there were many very small individuals and a few large ones. As the intensity of competition gradually increased, the degree of skewness increased as well. Decreased size – but increased skewness in size – is also seen to
be associated with increased density (and presumably competition) in cod (*Gadus morhua*) living off the coast of Norway (Figure 5.25).

More generally, we may also say that increased competition increased the degree of size inequality within the population, i.e. the extent to which total biomass was unevenly distributed amongst the different individuals (Weiner, 1990). Rather similar results have been obtained from a number of other populations of animals (Uchmanski, 1985) and plants (Uchmanski, 1985; Weiner & Thomas, 1986). Typically, populations experiencing the most intense competition have the greatest size inequality and often have a size distribution in which there are many small and a few large individuals. Characterizing a population by an arbitrary ‘average’ individual can obviously be very misleading under such circumstances, and can divert attention from the fact that intraspecific competition is a force affecting individuals, even though its effects may often be detected in whole populations.

### 5.10.2 Preempting resources

An indication of the way in which competition can exaggerate underlying inequalities in a population comes from observations on a natural, crowded population of the woodland annual *Impatiens pallida* in southeastern Pennsylvania. Over an 8-week period, growth was very much faster in large than in small plants – in fact, small plants did not grow at all (Figure 5.26a). This increased significantly the size inequality within the population (Figure 5.26b). Thus, the smaller a plant was initially, the more it was affected by neighbors. Plants that established early preempted or ‘captured’ space, and subsequently were little affected by intraspecific competition. Plants that emerged later entered a universe in which most of the available space had already been preempted; they were therefore greatly affected by intraspecific competition. Competition was asymmetric: there was a hierarchy. Some individuals were affected far more than others, and small initial differences were transformed by competition into much larger differences 8 weeks later.
If competition is asymmetric because superior competitors preempt resources, then competition is most likely to be asymmetric when it occurs for resources that are most liable to be preempted. Specifically, competition amongst plants for light, in which a superior competitor can overtop and shade an inferior, might be expected to lend itself far more readily to preemptive resource capture than competition for soil nutrients or water, where the roots of even a very inferior competitor will have more immediate access to at least some of the available resources than

Figure 5.25  (right) Values of skewness (in the frequency distribution of lengths) and density (a) and of skewness and mean length (b) are expressed as standard deviations from mean values for the years 1957–94 for cod (*Gadus morhua*) from the Skagerrak, off the coast of Norway. Despite marked fluctuations from year to year, much of it the result of variations in weather, skewness was clearly greatest at high densities (*r* = 0.58, *P* < 0.01) when lengths were smallest (*r* = −0.45, *P* < 0.05), that is, when competition was most intense. (After Lekve *et al.*, 2002.)

Figure 5.26  Asymmetric competition in a natural population of *Impatiens pallida*. (a) The increase in mass of survivors of different sizes over an 8-week period, and the distribution of initial sizes of those individuals that died over the same period. The horizontal axis is the same in each case. (b) The distribution of individual weights at the beginning (Gini coefficient, a measure of inequality, 0.39) and the end of this period (Gini coefficient, 0.48). (After Thomas & Weiner, 1989.)
the roots of its superiors. This expectation is borne out by the results of an experiment in which morning glory vines (*Ipomoea tricolor*) were grown as single plants in pots (‘no competition’), as several plants rooted in their own pots but with their stems intertwined on a single stake (‘shoots competing’), as several plants rooted in the same pot, but with their stems growing up their own stakes (‘roots competing’) and as several plants rooted in the same pot with their stems intertwined on one stake (‘shoots and roots competing’) (Figure 5.27). Despite the fact that root competition was more intense than shoot competition, in the sense that it led to a far greater decrease in the mean weight of individual plants, it was shoot competition for light that led to a much greater increase in size inequality.

Skewed distributions are one possible manifestation of hierarchical, asymmetric competition, but there are many others. For instance, Ziemba and Collins (1999) studied competition amongst larval salamanders (*Ambystoma tigrinum nebulosum*) that were either isolated or grouped together with competitors. The size of the largest surviving larvae was unaffected by competition (*P* = 0.42) but the smallest larvae were much smaller (*P* < 0.0001). This emphasizes that intraspecific competition is not only capable of exaggerating individual differences, it is also greatly affected by individual differences.

Asymmetric competition was observed on a much longer timescale in a population of the herbaceous perennial *Anemone hepatica* in Sweden (Figure 5.28) (Tamm, 1956). Despite the crops of seedlings that entered the population between 1943 and 1952, it is quite clear that the most important factor determining which individuals survived to 1956 was whether or not they were established in 1943. Of the 30 individuals that had reached large or intermediate size by 1943, 28 survived until 1956, and some of these had branched. By contrast, of the 112 plants that were either small in 1943 or appeared as seedlings subsequently, only 26 survived to 1956, and not one of these was sufficiently well established to have flowered. Similar patterns can be observed in tree populations. The survival rates, the birth rates and thus the fitnesses of the few established adults are high; those of the many seedlings and saplings are comparatively low.

These considerations illustrate a final, important general point: asymmetries tend to reinforce the regulatory powers of intraspecific competition. Tamm’s established plants were successful competitors year after year, but his small plants and seedlings were repeatedly unsuccessful. This guaranteed a near constancy in the number of established plants between 1943 and 1956. Each year there was a near-constant number of ‘winners’, accompanied by a variable number of ‘losers’ that not only failed to grow, but usually, in due course, died.

### 5.11 Territoriality

Territoriality is one particularly important and widespread phenomenon that results in asymmetric intraspecific competition. It occurs when there is active interference between individuals, such that a more or less exclusive area, the territory, is defended against intruders by a recognizable pattern of behavior.

Individuals of a territorial species that fail to obtain a territory often make no contribution whatsoever to future generations. Territoriality, then,
is a ‘contest’. There are winners (those that come to hold a territory) and losers (those that do not), and at any one time there can be only a limited number of winners. The exact number of territories (winners) is usually somewhat indeterminate in any one year, and certainly varies from year to year, depending on environmental conditions. Nevertheless, the contest nature of territoriality ensures, like asymmetric competition generally, a comparative constancy in the number of surviving, reproducing individuals. One important consequence of territoriality, therefore, is population regulation, or more particularly, the regulation of the number of territory holders. Thus, when territory owners die, or are experimentally removed, their places are often rapidly taken by newcomers. For instance, in great tit (Parus major) populations, vacated woodland territories are reoccupied by birds coming from hedgerows where reproductive success is noticeably lower (Krebs, 1971).

Some have felt that the regulatory consequences of territoriality must themselves be the root cause underlying the evolution of territorial behavior – territoriality being favored because the population as a whole benefitted from the rationing effects, which guaranteed that the population did not overexploit its resources (e.g. Wynne-Edwards, 1962). However, there are powerful and fundamental reasons for rejecting this ‘group selectionist’ explanation (essentially, it stretches evolutionary theory beyond reasonable limits): the ultimate cause of territoriality must be sought within the realms of natural selection, in some advantage accruing to the individual.

Any benefit that an individual does gain from territoriality, of course, must be set against the costs of defending the territory. In some animals this defense involves fierce combat between competitors, whilst in others there is a more subtle mutual recognition by competitors of one another’s keep-out signals (e.g. song or scent). Yet, even when the chances of physical injury are minimal, territorial animals typically expend energy in patrolling and advertising their territories, and these energetic costs must be exceeded by any benefits if territoriality is to be favored by natural selection (Davies & Houston, 1984; Adams, 2001).

Praw and Grant (1999), for example, investigated the costs and benefits to convict cichlid fish (Archocentrus nigrofasciatus) of defending food patches of different sizes. As patch size increased, the amount of food eaten by a patch defender increased (the benefit; Figure 5.29a), but the frequency of chasing intruders (the cost; Figure 5.29b) also increased. Evolution should favor an intermediate patch (territory) size at which the trade-off between costs and benefits is optimized, and indeed, the growth rate of defenders was greatest in intermediate-sized patches (Figure 5.29c).

On the other hand, explaining territoriality only in terms of a net benefit to the territory owner is rather like history always being written by the victors. There is another, possibly trickier question, which seems not to have been answered – could those individuals without a territory not do better by challenging the territory owners more often and with greater determination?

Of course, describing territoriality in terms of just ‘winners’ and ‘losers’ is an oversimplification. Generally, there are first, second and a range of consolation prizes – not all territories are equally valuable. This has been demonstrated in an unusually striking way in a study of oystercatchers (Haematopus ostralegus) on the Dutch coast, where pairs of birds defend both nesting territories on the salt marsh and feeding territories on the mudflats (Ens et al., 1992). For some birds (the ‘residents’), the feeding territory is simply an extension of the
nesting territory: they form one spatial unit. For other pairs, however (the ‘leapfrogs’), the nesting territory is further inland and hence separated spatially from the feeding territory (Figure 5.30a). Residents fledge many more offspring than do leapfrogs (Figure 5.30b), because they deliver far more food to them (Figure 5.30c). From an early age, resident chicks follow their parents onto the mudflats, taking each prey item as soon as it is captured. Leapfrog chicks, however, are imprisoned on their nesting territory prior to fledging; all their food has to be flown in. It is far better to have a resident than a leapfrog territory.

5.12 Self-thinning

We have seen throughout this chapter that intraspecific competition can influence the number of deaths, the number of births and the amount of growth within a population. We have illustrated this largely by looking at the end results of competition. But in practice, the effects are often progressive. As a cohort ages, the individuals grow in size, their requirements increase and they therefore compete at a greater and greater intensity. This in turn tends gradually to increase their risk of dying. But if some individuals die, then the density and the intensity of competition are decreased – which affects growth, which affects competition, which affects survival, which affects density, and so on.

5.12.1 Dynamic thinning lines

The patterns that emerge in growing, crowded cohorts of individuals were originally the focus of particular attention in plant populations. For example, perennial rye grass (Lolium perenne) was sown at a range of densities, and samples from each density were harvested after 14, 35, 76, 104 and 146 days (Figure 5.31a). Figure 5.31a has the same logarithmic axes – density and mean plant weight – as Figure 5.14. It is most important to appreciate the difference between the two. In Figure 5.14, each line represented a separate yield–density relationship at different ages of a cohort. Successive points along a line represent different initial sowing densities. In Figure 5.31, each line itself represents a different sowing density, and successive points along a line represent populations of this initial sowing density at different ages. The lines are therefore trajectories that follow a cohort through time. This is indicated by arrows, pointing from many small, young individuals (bottom right) to fewer, larger, older individuals (top left).

Mean plant weight (at a given age) was always greatest in the lowest density populations (Figure 5.31a). It is also clear that the highest density populations were the first to suffer substantial mortality. What is most noticeable, however, is that eventually, in all cohorts, density declined and mean plant weight increased in unison: populations progressed along roughly the same straight
The populations are said to have experienced self-thinning (i.e. a progressive decline in density in a population of growing individuals), and the line that they approached and then followed is known as a dynamic thinning line (Weller, 1990). The lower the sowing density, the later was the onset of self-thinning. In all cases, though, the populations initially followed a trajectory that was almost vertical, i.e. there was little mortality. Then, as they neared the thinning line, the populations suffered increasing amounts of mortality, so that the slopes of all the self-thinning trajectories gradually approached the dynamic thinning line and then progressed along it. Note also that Figure 5.31 has been drawn, following convention, with log density on the x-axis and log mean weight on the y-axis. This is not meant to imply that density is the independent variable on which mean weight depends. Indeed, it can be argued that mean weight increases naturally during plant growth, and this determines the decrease in density. The most satisfactory view is that density and mean weight are wholly interdependent.

Plant populations (if sown at sufficiently high densities) have repeatedly been found to approach and then follow a dynamic thinning line. For many years, all such lines were widely perceived as having a slope of roughly −3/2, and the relationship was often referred to as the ‘−3/2 power law’ (Yoda et al., 1963; Hutchings, 1983), since density (N) was seen as related to mean weight (w̄) by the equation:

\[ \log \bar{w} = \log c - \frac{3}{2} \log N \]  

(5.22)
or:
\[
\bar{w} = c N^{-1/2}
\]
(5.23)

where \(c\) is constant.

Note, however, that there are statistical problems in using Equations 5.22 and 5.23 to estimate the slope of the relationship (Weller, 1987). In particular, since \(\bar{w}\) is usually estimated as \(B/N\), where \(B\) is the total biomass per unit area, \(\bar{w}\) and \(N\) are inevitably correlated, and any relationship between them is, to a degree, spurious. It is therefore preferable to use the equivalent relationships, lacking autocorrelation:

\[\log B = \log c - \frac{1}{2} \log N\]
(5.24)

or:
\[B = c N^{-1/2}\].
(5.25)

5.12.2 Species and population boundary lines

In fact, in many cases where biomass–density relationships have been documented, it is not a single cohort that has been followed over time, but a series of crowded populations at different densities (and possibly different ages) that have been compared. In such cases, it is more correct to speak of a species boundary line – a line beyond which combinations of density and mean weight appear not to be possible for that species (Weller, 1990). Indeed, since what is possible for a species will vary with the environment in which it is living, the species boundary line will itself subsume a whole series of population boundary lines, each of which defines the limits of a particular population of that species in a particular environment (Sackville Hamilton et al., 1995).

Thus, a self-thinning population should approach and then track its population boundary line, which, as a trajectory, we would call its dynamic thinning line – but this need not also be its species boundary line. For example, the light regime, soil fertility, spatial arrangement of seedlings, and no doubt other factors may all alter the boundary line (and hence the dynamic thinning line) for a particular population (Weller, 1990; Sackville Hamilton et al., 1995). Soil fertility, for example, has been found in different studies to alter the slope of the thinning line, the intercept, neither, or both (Morris, 2002).

The influence of light is also worth considering in more detail, since it highlights a key feature of thinning and boundary lines. A slope of roughly \(-3/2\) means that mean plant weight is increasing faster than density is decreasing, and hence that total biomass is increasing (a slope of \(-1/2\) on a total biomass–density graph). But eventually this must stop: total biomass cannot increase indefinitely. Instead, the thinning line might be expected to change to a slope of \(-1\): that is, loss through mortality is exactly balanced by the growth of survivors, such that the total biomass remains constant (a horizontal line on a total biomass–density graph). This can be seen when populations of Lolium perenne (Figure 5.31b) were grown at low light intensities. A boundary (and thinning line) with a slope of \(-1\) was apparent.
at much lower densities than it would otherwise be. Clearly, the light regime can alter the population boundary line. This also emphasizes, however, that boundary lines with negative slopes steeper than $-1$ (whether or not they are exactly $-3/2$) imply limits to the allowable combinations of plant densities and mean weights that set in before the maximum biomass from an area of land has been reached. Possible reasons are discussed below.

5.12.3 A single boundary line for all species?

Intriguingly, when the thinning and boundary lines of all sorts of plants are plotted on the same figure, they all appear to have approximately the same slope and also to have intercepts (i.e. values of $c$ in Equation 5.24) falling within a narrow range (Figure 5.32). To the lower right of the figure are high-density populations of small plants (annual herbs and perennials with short-lived shoots), whilst to the upper left are sparse populations of very large plants, including coastal redwoods (Sequoia sempervirens), the tallest known trees. Fashions change in science as in everything else. At one time, ecologists looked at Figure 5.32 and saw uniformity – all plants marching in $-3/2$ time (e.g. White, 1980), with variations from the norm seen as either ‘noise’ or as only of minor interest. Subsequently, serious doubt was cast on the conformity of individual slopes to $-3/2$, and on the whole idea of a single, ideal thinning line (Weller, 1987, 1990; Zeide, 1987; Lonsdale, 1990). There really is no contradiction, though. On the one hand, the lines in Figure 5.32 occupy a very much smaller portion of the graph than one would expect by chance alone. There is apparently some fundamental phenomenon linking this whole spectrum of plant types: not an invariable ‘rule’ but an underlying trend. On the other hand, the variations between the lines are real and important and in as much need of explanation as the general trend.

5.12.4 A geometric basis for self-thinning

We proceed, therefore, by examining possible bases for the general trend, and then enquiring why different species or populations might display their own variations on this common theme. Two broad types of explanation for the trend have been proposed. The first (and for many years the only one) is geometric; the second is based on resource allocation in plants of different sizes.

The geometric argument runs as follows. In a growing cohort of plants, as the mass of the population increases, the leaf area index ($L$, the leaf area per unit area of land) does not keep on increasing. Instead, beyond a certain point, it remains constant irrespective of plant density ($N$). It is, in fact, precisely beyond this point that the population follows the dynamic thinning line. We can express this by writing:

$$L = \lambda N = \text{constant} \quad (5.26)$$

where $\lambda$ is the mean leaf area per surviving plant. However, the leaf area of individual plants increases as they grow, and so too therefore does their mean, $\lambda$. It is reasonable to expect $\lambda$, because it is an area, to be related to linear measurements of a plant, such as stem diameter, $D$, by a formula of the following type:

$$\lambda = aD^2 \quad (5.27)$$

where $a$ is a constant. Similarly, it is reasonable to expect mean plant weight, $\bar{w}$, to be related to $D$ by:

$$\bar{w} = bD^3 \quad (5.28)$$

where $b$ is also a constant. Putting Equations 5.26–5.28 together, we obtain:
\[ \tilde{w} = b(L/a)^{3/2} \cdot N^{-3/2} \]  
(5.29)

This is structurally equivalent to the \(-3/2\) power law in Equation 5.23, with the intercept constant, \(c\), given by \(b(L/a)^{3/2}\).

It is apparent, therefore, why thinning lines might generally be expected to have slopes of approximately \(-3/2\). Moreover, if the relationships in Equations 5.27 and 5.28 were roughly the same for all plant species, and if all plants supported roughly the same leaf area per unit area of ground \((L)\), then the constant \(c\) would be approximately the same for all species. On the other hand, suppose that \(L\) is not quite constant for some species (see Equation 5.26), or that the powers in Equations 5.27 and 5.28 are not exactly 2 or 3, or that the constants in these equations \((a\ and b)\) either vary between species or are not actually constants at all. Thinning lines will then have slopes that depart from \(-3/2\), and slopes and intercepts that vary from species to species. It is easy to see why, according to the geometric argument, there is a broad similarity in the behavior of different species, but also why, on closer examination, there are variations between species and no such thing as a single, ‘ideal’ thinning line.

Furthermore, contrary to the simple geometric argument, the yield–density relationship in a growing cohort need not depend only on the numbers that die and the way the survivors grow. We have seen (see Section 5.10) that competition is frequently highly asymmetric. If those that die in a cohort are predominantly the very smallest individuals, then density (individuals per unit area) will decline more rapidly as the cohort grows than it would otherwise do, and the slope will be shallower, especially in the early stages of self-thinning. This idea is supported by a comparison of self-thinning in normal Arabidopsis thaliana plants with self-thinning in mutants that overexpress phytochrome A, greatly reducing their shade tolerance, and making competition amongst them more asymmetric (Figure 5.33a).

It seems possible, too, to use departures from the assumptions built into Equations 5.26–5.29 to explain at least some of the variations from a ‘general’ \(-3/2\) rule. Osawa and Allen (1993) estimated a number of the parameters in these equations from data on the growth of individual plants of mountain beech (Nothofagus solandri) and red pine (Pinus densiflora). They estimated, for instance, that the exponents in Equations 5.27 and 5.28 were not 2 and 3, but 2.08 and 2.19 for mountain beech, and 1.63 and 2.41 for red pine. These suggest thinning slopes of \(-1.05\) in the first case and \(-1.48\) in the second, which compare quite remarkably well with the observed slopes of \(-1.06\) and \(-1.48\) (Figure 5.33b). The similarities between the estimates and observations for the intercept constants were equally impressive. These results show, therefore, that thinning lines with slopes other than \(-3/2\) can occur, but can be explicable in terms of the detailed biology of the species concerned – and that even when slopes of \(-3/2\) do occur, they may do so, as with red pine, for the ‘wrong’ reason \((-2.41/1.63\) rather than \(-3/2\)).

![Figure 5.33](image-url)
5.12.5 A resource-allocation basis for thinning boundaries

The increasing recognition that a variety of slope values might be expected, even on the geometric argument, along with the statistical difficulties of estimating slopes, has left the way open for alternative explanations for the underlying trend itself. Enquist et al. (1998) made use of the much more general model of West et al. (1997), which considered the most effective architectural designs of organisms (not just plants) for distributing acquired resources throughout those organisms. This suggested that the rate of resource supply per unit area and to find empirical support for this relationship.

But when the plants have arrived at an equilibrium with the rate of resource use per individual, $u$, should be related to mean plant weight, $\bar{w}$, according to the equation:

$$u = a\bar{w}^{4/3}$$  \hspace{1cm} (5.30)

where $a$ is a constant. Indeed, Enquist et al. (1998) were also able to find empirical support for this relationship.

They then argued that plants have evolved to make full use of the resources available, and so if $S$ is the rate of resource supply per unit area and $N_{\text{max}}$ the maximum allowable density of plants, then:

$$S = N_{\text{max}}u$$  \hspace{1cm} (5.31)

or, from Equation 5.30:

$$S = aN_{\text{max}}\bar{w}^{4/3}.$$  \hspace{1cm} (5.32)

But when the plants have arrived at an equilibrium with the rate of resource supply, $S$ should itself be constant. Hence:

$$\bar{w} = bN_{\text{max}}^{-4/3},$$  \hspace{1cm} (5.33)

where $b$ is another constant. In short, the expected slope of a population boundary on this argument is $-4/3$ rather than $-3/2$. Enquist and colleagues themselves considered the available data to be more supportive of their prediction of a slope of $-4/3$ than the more conventional $-3/2$. This has not, however, been the conclusion drawn either from previous data surveys or from the analysis of subsequent experiments (e.g. Figure 5.33a; Stoll et al., 2002). In part, the discrepancy may have arisen because the geometric argument is focused on light acquisition, and the data collected to test it have likewise been focused on above-ground plant parts (photosynthetic or support tissue); whereas Enquist et al.’s is a much more general resource-acquisition argument, and at least some of their data were based on overall plant weights (leaves, shoots and roots). Related to this, Enquist et al.’s data sets were focused on maximum densities of large numbers of species, whereas other analyses have focused on the self-thinning process, which occurs largely before the overall resource-determined limit has been reached. Again, therefore, there may be no contradiction between the two approaches.

5.12.6 Self-thinning in animal populations

Animals, whether they are sessile or mobile, must also ‘self-thin’, insofar as growing individuals within a cohort increasingly compete with one another and reduce their own density. There is nothing linking all animals quite like the shared need for light interception that links all plants, so there is even less likelihood of a general self-thinning ‘law’ for animals. On the other hand, crowded sessile animals can, like plants, be seen as needing to pack ‘volumes’ beneath an approximately constant area, and mussels, for example, have been found to follow a thinning line with a slope of $-1.4$, and barnacles a line with a slope of $-1.6$ (Hughes & Griffiths, 1988). Moreover, self-thinning in the gregarious tunicate, *Pyura praeputialis*, on the coast of Chile was found to follow a slope of only $-1.2$; but when the analysis was modified to acknowledge that rocky shore invertebrates are more ‘three-dimensional’ than plants, and may fit more than one layer into a fully occupied area (as opposed to the constant leaf area index of plants), then the estimated slope was $-1.5$ (Figure 5.34a).

For mobile animals, it has been suggested that the relationship between metabolic rate and body size could generate thinning lines with slopes of $-4/3$ (Begon et al., 1986). However, the generality of this is probably even more questionable than the ‘rules’ in plants, given variations in resource supply, variations in the coefficients in the underlying relationships, and the possibilities of self-thinning depending on, say, territorial behavior rather than simply food availability (Steingrimsson & Grant, 1999). Nonetheless, evidence of self-thinning in animals is increasingly reported, especially in fish, even if the basis for it remains uncertain (e.g. Figure 5.34b).

Plants are not so consistent in their pattern of self-thinning as was once thought. It may be that animals are not much less bound than plants by ‘general’ self-thinning rules.

**Summary**

Intraspecific competition is defined and explained. Exploitation and interference are distinguished, and the commonly one-sided nature of competition is emphasized.

We describe the effects of intraspecific competition on rates of mortality and fecundity, distinguishing under-, over- and exactly compensating density dependence. We explain, however, that density itself is usually just a convenient expression of crowding or shortage of resources.
These effects at the individual level lead in turn to patterns, and regulatory tendencies, at the population level. The carrying capacity is defined and its limitations are explained, along with the domed nature of net recruitment curves and the sigmoidal nature of population growth curves.

We describe the effects of intraspecific competition on rates of growth, explaining the ‘law of constant final yield’, especially in modular organisms.

The use of $k$ values in quantifying intraspecific competition is described, and scramble and contest competition are distinguished.

We introduce the use of mathematical models in ecology generally, then go on to develop a model of a population with discrete breeding seasons subject to intraspecific competition. The model illustrates the tendency of time lags to provoke population fluctuations and that different types of competition may lead to different types of population dynamics, including patterns of deterministic chaos – the nature and importance of which are themselves explained. A model with continuous breeding is also developed, leading to the logistic equation.

The importance of individual differences in generating asymmetries in competition is explained, as is the importance of competition in generating individual differences. Asymmetries tend to enhance regulation; territoriality is a particularly important example of this.

The progressive effects of competition on growth and mortality may often be interlinked in the process of self-thinning, which has been a particular focus in plant populations. We explain the nature of dynamic thinning lines and the $-3/2$ power law when single cohorts are followed, and also species and population boundary lines when a series of crowded populations is observed at different densities. We address the question of whether there is a single boundary line for all species.

We explain how two broad types of explanation for the consistent trend amongst species have been proposed: those based on geometry and those based on resource allocation in plants of different sizes.

Finally we examine self-thinning in animal populations and conclude that plants are not so consistent in their pattern of self-thinning as was once thought, while animals are not much less bound than plants by ‘general’ self-thinning rules.

Figure 5.34 (a) Self-thinning in the gregarious tunicate, Pyura praecutialis, where density has been modified to include an ‘effective area’ which incorporates the number of layers in the tunicate colonies. The estimated slope is $-1.49$ (95% CI $-1.59$ to $-1.39$, $P < 0.001$). (After Guíñez & Castilla, 2001.) (b) Dynamic thinning lines for 23 year-classes of sea trout, Salmo trutta, from an English Lake District stream, with the position of the mean regression line (slope $= -1.35$) indicated by the arrows (After Elliott, 1993.)
6.1 Introduction

All organisms in nature are where we find them because they have moved there. This is true for even the most apparently sedentary of organisms, such as oysters and redwood trees. Their movements range from the passive transport that affects many plant seeds to the apparently purposeful actions of many mobile animals. Dispersal and migration are used to describe aspects of the movement of organisms. The terms are defined for groups of organisms, although it is of course the individual that moves.

Dispersal is most often taken to mean a spreading of individuals away from others, and is therefore an appropriate description for several kinds of movements: (i) of plant seeds or starfish larvae away from each other and their parents; (ii) of voles from one area of grassland to another, usually leaving residents behind and being counterbalanced by the dispersal of other voles in the other direction; and (iii) of land birds amongst an archipelago of islands (or aphids amongst a mixed stand of plants) in the search for a suitable habitat.

Migration is most often taken to mean the mass directional movements of large numbers of a species from one location to another. The term therefore applies to classic migrations (the movements of locust swarms, the intercontinental journeys of birds) but also to less obvious examples like the to and fro movements of shore animals following the tidal cycle. Whatever the precise details of dispersal in particular cases, it will be useful in this chapter to divide the process into three phases: starting, moving and stopping (South et al., 2002) or, put another way, emigration, transfer and immigration (Ims & Yoccoz, 1997). The three phases differ (and the questions we ask about them differ) both from a behavioral point of view (what triggers the initiation and cessation of movement?, etc.) and from a demographic point of view (the distinction between loss and gain of individuals, etc.).

The division into these phases also emphasizes that dispersal can refer to the process by which individuals, in leaving, escape from the immediate environment of their parents and neighbors; but it can also often involve a large element of discovery or even exploration. It is useful, too, to distinguish between natal dispersal and breeding dispersal (Clobert et al., 2001). The former refers to the movement between the natal area (i.e. where the individual was born) and where breeding first takes place. This is the only type of dispersal possible in a plant. Breeding dispersal is movement between two successive breeding areas.

6.2 Active and passive dispersal

Like most biological categories, the distinction between active and passive dispersers is blurred at the edges. Passive dispersal in air currents, for example, is not restricted to plants. Young spiders that climb to high places and then release a gossamer thread that carries them on the wind are then passively at the mercy of air currents; i.e. ‘starting’ is active but moving itself is effectively passive. Even the wings of insects are often simply aids to what is effectively passive movement (Figure 6.1).

6.2.1 Passive dispersal: the seed rain

Most seeds fall close to the parent and their density declines with distance from that parent. This is the case for wind-dispersed seeds and also for those that are ejected actively by maternal tissue (e.g. many legumes). The eventual destination of the dispersed offspring is determined by the original location of the parent and by the relationship relating disperser density to distance from parent, but the detailed microhabitat of that destination is left to chance. Dispersal is nonexploratory; discovery is a matter of chance. Some animals have essentially this same type of dispersal. For
example, the dispersal of most pond-dwelling organisms without a free-flying stage depends on resistant wind-blown structures (e.g. gemmules of sponges, cysts of brine shrimps).

The density of seeds is often low immediately under the parent, rises to a peak close by and then falls off steeply with distance (Figure 6.2a). However, there are immense practical problems in studying seed dispersal (i.e. in following the seeds), and these become increasingly irresolvable further from the source. Greene and Calogeropoulos (2001) liken any assertion that ‘most seeds travel short distances’ to a claim that most lost keys and contact lenses fall close to streetlights. Certainly, the very few studies of long-distance dispersal that have been carried out suggest that seed density declines only very slowly at larger distances from the original source (Figure 6.2b), and even a few long-distance dispersers may be crucial in either invasion or recolonization dispersal (see Section 6.3.1).

6.2.2 Passive dispersal by a mutualistic agent

Uncertainty of destination may be reduced if an active agent of dispersal is involved. The seeds of many herbs of the woodland...
floor have spines or prickles that increase their chance of being carried passively on the coats of animals. The seeds may then be concentrated in nests or burrows when the animal grooms itself. The fruits of many shrubs and lower canopy trees are fleshy and attractive to birds, and the seed coats resist digestion in the gut. Where the seed is dispersed to is then somewhat less certain, depending on the defecating behavior of the bird. It is usually presumed that such associations are ‘mutualistic’ (beneficial to both parties – see Chapter 13): the seed is dispersed in a more or less predictable fashion; the disperser consumes either the fleshy ‘reward’ or a proportion of the seeds (those that it finds again).

There are also important examples in which animals are dispersed by an active agent. For instance, there are many species of mite that are taken very effectively and directly from dung pat to dung pat, or from one piece of carrion to another, by attaching themselves to dung or carrion beetles. They usually attach to a newly emerging adult, and leave again when that adult reaches a new patch of dung or carrion. This, too, is typically mutualistic: the mites gain a dispersive agent, and many of them attack and eat the eggs of flies that would otherwise compete with the beetles.

6.2.3 Active discovery and exploration

Many other animals cannot be said to explore, but they certainly control their settlement (‘stopping’, see Section 6.1.1) and cease movement only when an acceptable site has been found. For example, most aphids, even in their winged form, have powers of flight that are too weak to counteract the forces of prevailing winds. But they control their take-off from their site of origin, they control when they drop out of the windstream, and they make additional, often small-scale flights if their original site of settlement is unsatisfactory. In a precisely analogous manner, the larvae of many river invertebrates make use of the flowing column of water for dispersing from hatching sites to appropriate microhabitats (‘invertebrate drift’) (Brittain & Eikeland, 1988). The dispersal of aphids in the wind and of drifting invertebrates in streams, therefore, involves ‘discovery’, over which they have some, albeit limited, control.

Other animals explore, visiting many sites before returning to a favored suitable one. For example, in contrast to their drifting larvae, most adults of freshwater insects depend on flight for upstream dispersal and movement from stream to stream. They
explore and, if successful, discover, suitable sites within which to lay their eggs: starting, moving and stopping are all under active control.

6.2.4 Clonal dispersal

In almost all modular organisms (see Section 4.2.1), an individual genet branches and spreads its parts around it as it grows. There is a sense, therefore, in which a developing tree or coral actively disperses its modules into, and explores, the surrounding environment. The interconnections of such a clone often decay, so that it becomes represented by a number of dispersed parts. This may result ultimately in the product of one zygote being represented by a clone of great age that is spread over great distances. Some clones of the rhizomatous bracken fern (*Pteridium aquilinum*) were estimated to be more than 1400 years old and one extended over an area of nearly 14 ha (Oinonen, 1967).

We can recognize two extremes in a continuum of strategies in clonal dispersal (Lovett Doust & Lovett Doust, 1982; Sackville Hamilton et al., 1987). At one extreme, the connections between modules are long and the modules themselves are widely spaced. These have been called ‘guerrilla’ forms, because they give the plant, hydroid or coral a character like that of a guerrilla army. Fugitive and opportunist, they are constantly on the move, disappearing from some territories and penetrating into others. At the other extreme are ‘phalanx’ forms, named by analogy with the phalanxes of a Roman army, tightly packed with their shields held around them. Here, the connections are short and the modules are tightly packed, and the organisms expand their clones slowly, retain their original site occupancy for long periods, and neither penetrate readily amongst neighboring plants nor are easily penetrated by them.

Even amongst trees, it is easy to see that the way in which the buds are placed gives them a guerrilla or a phalanx type of growth form. The dense packing of shoot modules in species like cypresses (*Cupressus*) produces a relatively undispersed and impenetrable phalanx canopy, whilst many loose-structured, broad-leaved trees (*Acacia, Betula*) can be seen as guerrilla canopies, bearing buds that are widely dispersed and shoots that interweave with the buds and branches of neighbors. The twining or clambering lianas in a forest are guerrilla growth forms *par excellence*, dispersing their foliage and buds over immense distances, both vertically and laterally.

The way in which modular organisms disperse and display their modules affects the ways in which they interact with their neighbors. Those with a guerrilla form will continually meet and compete with other species and other genets of their own kind. With a phalanx structure, however, most meetings will be between modules of a single genet. For a tussock grass or a cypress tree, competition must occur very largely between parts of itself.

Clonal growth is most effective as a means of dispersal in aquatic environments. Many aquatic plants fragment easily, and the parts of a single clone become independently dispersed because they are not dependent on the presence of roots to maintain their water relations. The major aquatic weed problems of the world are caused by plants that multiply as clones and fragment and fall to pieces as they grow: duckweeds (*Lemna* spp.), the water hyacinth (*Eichhornia crassipes*), Canadian pond weed (*Elodea Canadensis*) and the water fern *Salvinia*.

6.3 Patterns of distribution: dispersion

The movements of organisms affect the spatial pattern of their distribution (their *dispersion*) and we can recognize three main patterns of dispersion, although they too form part of a continuum (Figure 6.3).

*Random* dispersion occurs when there is an equal probability of an organism occupying any point in space (irrespective of the position of any others). The result is that individuals are unevenly distributed because of chance events.

*Regular* dispersion (also called a *uniform* or *even* distribution or *overdispersion*) occurs either when an individual has a tendency to avoid other individuals, or when individuals that are especially close to others die. The result is that individuals are more evenly spaced than expected by chance.

*Aggregated* dispersion (also called a *contagious* or *clumped* distribution or *underdispersion*) occurs either when individuals tend to be attracted to (or are more likely to survive in) particular parts of the environment, or when the presence of one individual

![Figure 6.3](image)
attracts, or gives rise to, another close to it. The result is that individuals are closer together than expected by chance.

How these patterns appear to an observer, however, and their relevance to the life of other organisms, depends on the spatial scale at which they are viewed. Consider the distribution of an aphid living on a particular species of tree in a woodland. At a large scale, the aphids will appear to be aggregated in particular parts of the world, i.e. in woodlands as opposed to other types of habitat. If samples are smaller and taken only in woodlands, the aphids will still appear to be aggregated, but now on their host tree species rather than on trees in general. However, if samples are smaller still (25 cm², about the size of a leaf) and are taken within the canopy of a single tree, the aphids might appear to be randomly distributed over the tree as a whole. At an even smaller scale (c. 1 cm²) we might detect a regular distribution because individual aphids on a leaf avoid one another.

### 6.3.1 Patchiness

In practice, the populations of all species are patchily distributed at some scale or another, but it is crucial to describe dispersion at scales that are relevant to the lifestyle of the organisms concerned. MacArthur and Levins (1964) introduced the concept of environmental grain to make this point. For example, the canopy of an oak–hickory forest, from the point of view of a bird like the scarlet tanager (*Piranga olivacea*) that forages indiscriminately in both oaks and hickories, is fine grained: i.e. it is patchy, but the birds experience the habitat as an oak–hickory mixture. The habitat is coarse grained, however, for defoliating insects that attack either oaks or hickories preferentially: they experience the habitat one patch at a time, moving from one preferred patch to another (Figure 6.4).

Patchiness may be a feature of the physical environment: islands surrounded by water, rocky outcrops in a moorland, and so on. Equally important, patchiness may be created by the activities of organisms themselves; by their grazing, the deposition of dung, trampling or by the local depletion of water and mineral resources. Patches in the environment that are created by the activity of organisms have life histories. A gap created in a forest by a falling tree is colonized and grows up to contain mature trees, whilst other trees fall and create new gaps. The dead leaf in a grassland area is a patch for colonization by a succession of fungi and bacteria that eventually exhaust it as a resource, but new dead leaves arise and are colonized elsewhere.

Patchiness, dispersal and scale are tied intimately together. A framework that is useful in thinking about this distinguishes between local and landscape scales (though what is ‘local’ to a worm is very different from what is local to the bird that eats it) and between turnover and invasion dispersal (Bullock et al., 2002). Turnover dispersal at the local scale describes the movement into a gap from occupied habitat immediately surrounding the gap; whereas that gap may also be invaded or colonized by individuals moving in from elsewhere in the surrounding community. At the landscape scale, similarly, dispersal may be part of an on-going turnover of extinction and recolonization of occupiable patches within an otherwise unsuitable habitat matrix (e.g. islands in a stream: ‘metapopulation dynamics’ – see Section 6.9, below), or dispersal may result in the invasion of habitat by a ‘new’ species expanding its range.

### 6.3.2 Forces favoring aggregations (in space and time)

The simplest evolutionary explanation for the patchiness of populations is that organisms aggregate when and where they find resources and conditions that favor reproduction and survival. These resources and conditions are usually patchily distributed in both space and time. It pays (and has paid in evolutionary time)
to disperse to these patches when and where they occur. There are, however, other specific ways in which organisms may gain from being close to neighbors in space and time.

An elegant theory identifying a selective advantage to individuals that aggregate with others was suggested by Hamilton (1971) in his paper ‘Geometry for the selfish herd’. He argued that the risk to an individual from a predator may be lessened if it places another potential prey individual between itself and the predator. The consequence of many individuals doing this is bound to be an aggregation. The ‘domain of danger’ for individuals in a herd is at the edge, so that an individual would gain an advantage if its social status allowed it to assimilate into the center of a herd. Subordinate individuals might then be forced into the regions of greater danger on the edge of the flock. This seems to be the case in reindeer \((Rangifer tarandus)\) and wood pigeons \((Columba palumbus)\), where a newcomer may have to join the herd or flock at its risky perimeter and can only establish itself in a more protected position within the flock after social interaction (Murton et al., 1966).

Individuals may also gain from living in groups if this helps to locate food, give warning of predators or if it pays for individuals to join forces in fighting off a predator (Pulliam & Caraco, 1984). The principle of the selfish herd as described for the aggregation of organisms in space is just as appropriate for the synchronous appearance of organisms in time. The individual that is precocious or delayed in its appearance, outside the norm for its population, may be at greater risk from predators than those conformist individuals that take part in ‘flooding the market’ thereby diluting their own risk. Amongst the most remarkable examples of synchrony are the ‘periodic cicadas’ (insects), the adults of which emerge simultaneously after 13 or 17 years of life underground as nymphs. Williams et al. (1993) studied the mortality of populations of 13-year periodic cicadas that emerged in northwestern Arkansas in 1985. Birds consumed almost all of the standing crop of cicadas when the density was low, but only 15–40% when the cicadas reached peak density. Predation then rose to near 100% as the cicada density fell again (Figure 6.5). Equivalent arguments apply to the many species of tree, especially in temperate regions, that have synchronous ‘mast’ years (see Section 9.4).

### 6.3.3 Forces diluting aggregations: density-dependent dispersal

There are also strong selective pressures that can act against aggregation in space or time. In some species a group of individuals may actually concentrate a predator’s attention (the opposite effect to the ‘selfish herd’). However, the foremost diluting forces are certain to be the more intense competition suffered by crowded individuals (see Chapter 5) and the direct interference between such individuals even in the absence of a shortage of resources. One likely consequence is that the highest rates of dispersal will be away from the most crowded patches: density-dependent emigration dispersal (Figure 6.6) (Sutherland et al., 2002), though as we shall see below, density-dependent dispersal is by no means a general rule.

Overall, though, the types of distribution over available patches found in nature are bound to be compromises between forces attracting individuals to disperse towards one another and forces provoking individuals to disperse away from one another. As we shall see in a later chapter, such compromises are conventionally crystallized in the ‘ideal free’ and other theoretical distributions (see Section 9.6.3).

### 6.4 Patterns of migration

#### 6.4.1 Tidal, diurnal and seasonal movements

Individuals of many species move \textit{en masse} from one habitat to another and back again repeatedly during their life. The timescale involved may be hours, days, months or years. In some cases, these movements have the effect of maintaining the
organism in the same type of environment. This is the case in the movement of crabs on a shoreline: they move with the advance and retreat of the tide. In other cases, diurnal migration may involve moving between two environments: the fundamental niches of these species can only be satisfied by alternating life in two distinct habitats within each day of their lives. For example, some planktonic algae both in the sea and in lakes descend to the depths at night but move to the surface during the day. They accumulate phosphorus and perhaps other nutrients in the deeper water at night before returning to photosynthesize near the surface during daylight hours (Salonen et al., 1984). Other species aggregate into tight populations during a resting period and separate from each other when feeding. For example, most land snails rest in confined humid microhabitats by day, but range widely when they search for food by night.

Many organisms make seasonal migrations – again, either tracking a favorable habitat or benefitting from different, complementary habitats. The altitudinal migration of grazing animals in mountainous regions is one example. The American elk (Cervus elaphus) and mule deer (Odocoileus hemionus), for instance, move up into high mountain areas in the summer when food supplies become abundant during the warm summer period, and move south to savannas in the fall when food becomes abundant only after the rainy season. Both are areas in which seasons of comparative glut and famine alternate. Migrants then make a large contribution to the diversity of a local fauna. Of the 589 species of birds (excluding seabirds) that breed in the Palaearctic (temperate Europe and Asia), 40% spend the winter elsewhere (Moreau, 1952). Of those species that leave for the winter, 98% travel south to Africa. On an even larger scale, the Arctic tern (Sterna paradisaea) travels from its Arctic breeding ground to the Antarctic pack ice and back each year – about 10,000 miles (16,100 km) each way (although unlike many other migrants it can feed on its journey).

The same species may behave in different ways in different places. All robins (Erithacus rubecula) leave Finland and Sweden in winter, but on the Canary Islands the species is resident the whole year-round. In most of the intervening countries, a part of the population migrates and a part remains resident. Such variations are in some cases associated with clear evolutionary divergence. This is true of the knot (Calidris canutus), a species of small wading bird mostly breeding in remote areas of the Arctic tundras and ‘wintering’ in the summers of the southern hemisphere. At least five subspecies appear to have diverged in the Late Pleistocene (based on genetic evidence from the sequencing of mitochondrial DNA), and these now have strikingly different patterns of distribution and migration (Figure 6.7).
Long-distance migration is a feature of many other groups too. Baleen whales in the southern hemisphere move south in summer to feed in the food-rich waters of the Antarctic. In winter they move north to breed (but scarcely to feed) in tropical and subtropical waters. Caribou (Rangifer tarandus) travel several hundred kilometers per year from northern forests to the tundra and back. In all of these examples, an individual of the migrating species may make the return journey several times.

Many long-distance migrants, however, make only one return journey during their lifetime. They are born in one habitat, make their major growth in another habitat, but then return to breed and die in the home of their infancy. Eels and migratory salmon provide classic examples. The European eel (Anguilla anguilla) travels from European rivers, ponds and lakes across the Atlantic to the Sargasso Sea, where it is thought to reproduce and die (although spawning adults and eggs have never actually been caught there). The American eel (Anguilla rostrata) makes a comparable journey from areas ranging between the Guianas in the south, to southwest Greenland in the north. Salmon make a comparable transition, but from a freshwater egg and juvenile phase to mature as a marine adult. The fish then returns to freshwater sites to lay eggs. After spawning, all Pacific salmon (Oncorhynchus nerka) die without ever returning to the sea. Many Atlantic salmon (Salmo salar) also die after spawning, but some survive to return to the sea and then migrate back upstream to spawn again.

6.4.3 ‘One-way only’ migration

In some migratory species, the journey for an individual is on a strictly one-way ticket. In Europe, the clouded yellow (Colias croceus), red admiral (Vanessa atalanta) and painted lady (Vanessa cardui) butterflies breed at both ends of their migrations. The individuals that reach Great Britain in the summer breed there, and their offspring fly south in autumn and breed in the Mediterranean region – the offspring of these in turn come north in the following summer.

Most migrations occur seasonally in the life of individuals or of populations. They usually seem to be triggered by some
external seasonal phenomenon (e.g. changing day length), and sometimes also by an internal physiological clock. They are often preceded by quite profound physiological changes such as the accumulation of body fat. They represent strategies evolved in environments where seasonal events like rainfall and temperature cycles are reliably repeated from year to year. There is, however, a type of migration that is tactical, forced by events such as overcrowding, and appears to have no cycle or regularity. These are most common in environments where rainfall is not seasonally reliable. The economically disastrous migration plagues of locusts in arid and semiarid regions are the most striking examples.

6.5 Dormancy: migration in time

An organism gains in fitness by dispersing its progeny as long as the progeny are more likely to leave descendants than if they remained undispersed. Similarly, an organism gains in fitness by delaying its arrival on the scene, so long as the delay increases its chances of leaving descendants. This will often be the case when conditions in the future are likely to be better than those in the present. Thus, a delay in the recruitment of an individual to a population may be regarded as ‘migration in time’.

Organisms generally spend their period of delay in a state of dormancy. This relatively inactive state has the benefit of conserving energy, which can then be used during the period following the delay. In addition, the dormant phase of an organism is often more tolerant of the adverse environmental conditions prevailing during the delay (i.e. tolerant of drought, extremes of temperature, lack of light and so on). Dormancy can be either predictive or consequential (Müller, 1970). Predictive dormancy is initiated in advance of the adverse conditions, and is most often found in predictable, seasonal environments. It is generally referred to as ‘diapause’ in animals, and in plants as ‘innate’ or ‘primary’ dormancy (Harper, 1977). Consequential (or ‘secondary’) dormancy, on the other hand, is initiated in response to the adverse conditions themselves.

6.5.1 Dormancy in animals: diapause

Diapause has been most intensively studied in insects, where examples occur in all developmental stages. The common field grasshopper Chorthippus brunneus is a fairly typical example. This annual species passes through an obligatory diapause in its egg stage, where, in a state of arrested development, it is resistant to the cold winter conditions that would quickly kill the nymphs and adults. In fact, the eggs require a long cold period before development can start again (around 5 weeks at 0°C, or rather longer at a slightly higher temperature) (Richards & Waloff, 1954). This ensures that the eggs are not affected by a short, freak period of warm winter weather that might then be followed by normal, dangerous, cold conditions. It also means that there is an enhanced synchronization of subsequent development in the population as a whole. The grasshoppers ‘migrate in time’ from late summer to the following spring.

Diapause is also common in species with more than one generation per year. For instance, the fruit-fly Drosophila obscura passes through four generations per year in England, but enters diapause during only one of them (Begon, 1976). This facultative diapause shares important features with obligatory diapause: it enhances survivorship during a predictably adverse winter period, and it is experienced by resistant diapause adults with arrested gonadal development and large reserves of stored abdominal fat. In this case, synchronization is achieved not only during diapause but also prior to it. Emerging adults react to the short daylengths of the fall by laying down fat and entering the diapause state; they recommence development in response to the longer days of spring. Thus, by relying, like many species, on the utterly predictable photoperiod as a cue for seasonal development, D. obscura enters a state of predictive diapause that is confined to those generations that inevitably pass through the adverse conditions.

Consequential dormancy may be expected to evolve in environments that are relatively unpredictable. In such circumstances, there will be a disadvantage in responding to adverse conditions only after they have appeared, but this may be outweighed by the advantages of: (i) responding to favorable conditions immediately after they reappear; and (ii) entering a dormant state only if adverse conditions do appear. Thus, when many mammals enter hibernation, they do so (after an obligatory preparatory phase) in direct response to the adverse conditions. Having achieved ‘resistance’ by virtue of the energy they conserve at a lowered body temperature, and having periodically emerged and monitored their environment, they eventually cease hibernation whenever the adversity disappears.

6.5.2 Dormancy in plants

Seed dormancy is an extremely widespread phenomenon in flowering plants. The young embryo ceases development whilst still attached to the mother plant and enters a phase of suspended activity, usually losing much of its water and becoming dormant in a desiccated condition. In a few species of higher plants, such as some mangroves, a dormant period is absent, but this is very much the exception – almost all seeds are dormant when they are shed from the parent and require special stimuli to return them to an active state (germination).

Dormancy in plants, though, is not confined to seeds. For example, as the sand sedge Carex arenaria grows, it tends to accumulate dormant buds along the length of its predominantly linear rhizome. These may remain alive but dormant long after
the shoots with which they were produced have died, and they have been found in numbers of up to 400–500 m\(^2\) (Noble et al., 1979). They play a role analogous to the bank of dormant seeds produced by other species.

Indeed, the very widespread habit of deciduousness is a form of dormancy displayed by many perennial trees and shrubs. Established individuals pass through periods, usually of low temperatures and low light levels, in a leafless state of low metabolic activity.

Three types of dormancy have been distinguished.

1 **Innate dormancy** is a state in which there is an absolute requirement for some special external stimulus to reactivate the process of growth and development. The stimulus may be the presence of water, low temperature, light, photoperiod or an appropriate balance of near- and far-red radiation. Seedlings of such species tend to appear in sudden flushes of almost simultaneous germination. Deciduousness is also an example of innate dormancy.

2 **Enforced dormancy** is a state imposed by external conditions (i.e. it is consequential dormancy). For example, the Missouri goldenrod *Solidago missouriensis* enters a dormant state when attacked by the beetle *Trirhabda canadensis*. Eight clones, identified by genetic markers, were followed prior to, during and after a period of severe defoliation. The clones, which varied in extent from 60 to 350 m\(^2\) and from 700 to 20,000 rhizomes, failed to produce any above-ground growth (i.e. they were dormant) in the season following defoliation and had apparently died, but they reappeared 1–10 years after they had disappeared, and six of the eight bounced back strongly within a single season (Figure 6.8). Generally, the progeny of a single plant with enforced dormancy may be dispersed in time over years, decades or even centuries. Seeds of *Chenopodium album* collected from archeological excavations have been shown to be viable when 1700 years old (Ødum, 1965).

3 **Induced dormancy** is a state produced in a seed during a period of enforced dormancy in which it acquires some new requirement before it can germinate. The seeds of many agricultural and horticultural weeds will germinate without a light stimulus when they are released from the parent; but after a period of enforced dormancy they require exposure to light before they will germinate. For a long time it was a puzzle that soil samples taken from the field to the laboratory would quickly generate huge crops of seedlings, although these same seeds had failed to germinate in the field. It was a simple idea of genius that prompted Wesson and Wareing (1969) to collect soil samples from the field at night and bring them to the laboratory in darkness. They obtained large crops of seedlings from the soil only when the samples were exposed to light. This type of induced dormancy is responsible for the accumulation of large populations of seeds in the soil. In nature they germinate only when they are brought to the soil surface by earthworms or other burrowing animals, or by the exposure of soil after a tree falls.

Seed dormancy may be induced by radiation that contains a relatively high ratio of far-red (730 nm) to near-red (approximately 660 nm) wavelengths, a spectral composition characteristic of light that has filtered through a leafy canopy. In nature, this must have the effect of holding sensitive seeds in the dormant state when they land on the ground under a canopy, whilst releasing them into germination only when the overshadowing plants have died away.

Most of the species of plants with seeds that persist for long in the soil are annuals and biennials, and they are mainly weedy species – opportunists waiting (literally) for an opening. They largely lack features that will disperse them extensively in space. The seeds of trees, by contrast, usually have a very short expectation of life in the soil, and many are extremely difficult to store artificially for more than 1 year. The seeds of many tropical trees are particularly short lived: a matter of weeks or even days. Amongst trees,
the most striking longevity is seen in those that retain the seeds in cones or pods on the tree until they are released after fire (many species of *Eucalyptus* and *Pinus*). This phenomenon of serotiny protects the seeds against risks on the ground until fire creates an environment suitable for their rapid establishment.

### 6.6 Dispersal and density

Density-dependent emigration was identified in Section 6.3.3 as a frequent response to overcrowding. We turn now to the more general issue of the density dependence of dispersal and also to the evolutionary forces that may have led to any density dependences that are apparent. In doing so, it is important to bear in mind the point made earlier (see Section 6.1.1): that ‘effective’ dispersal (from one place to another) requires emigration, transfer and immigration. The density dependences of the three need not be the same.

#### 6.6.1 Inbreeding and outbreeding

Much of this chapter is devoted to the demographic or ecological consequences of dispersal, but there are also important genetic and evolutionary consequences. Any evolutionary ‘consequence’ is, of course, then a potentially important selective force favoring particular patterns of dispersal or indeed the tendency to disperse at all. In particular, when closely related individuals breed, their offspring are likely to suffer an ‘inbreeding depression’ in fitness (Charlesworth & Charlesworth, 1987), especially as a result of the expression in the phenotype of recessive deleterious alleles. With limited dispersal, inbreeding becomes more likely, and inbreeding avoidance is thus a force favoring dispersal. On the other hand, many species show local adaptation to their immediate environment (see Section 1.2). Longer distance dispersal may therefore bring together genotypes adapted to different local environments, which on mating give rise to low-fitness offspring adapted to neither habitat. This is called ‘outbreeding depression’, resulting from the break-up of coadapted combinations of genes – a force acting against dispersal. The situation is complicated by the fact that inbreeding depression is most likely amongst populations that normally outbreed, since inbreeding itself will purge populations of their deleterious recessives. None the less, natural selection can be expected to favor a pattern of dispersal that is in some sense intermediate – maximizing fitness by avoiding both inbreeding and outbreeding depression, though these will clearly be by no means the only selective forces acting on dispersal.

Certainly, there are several examples in plants of inbreeding and outbreeding depression when pollen is transferred from either close or distant donors, and in some cases both effects can be demonstrated in a single experiment. For example, when larkspur (*Delphinium nelsonii*) offspring were generated by hand pollinating with pollen brought from 1, 3, 10 and 30 m to the receptor flowers (Figure 6.9), both inbreeding and outbreeding depression in fitness were apparent.

#### 6.6.2 Avoiding kin competition

In fact, inbreeding avoidance is not the only force likely to favor natal dispersal of offspring away from their close relatives. Such
dispersal will also be favored because it decreases the likelihood of competitive effects being directed at close kin. This was explained in a classic modeling paper by Hamilton and May (1977; see also Gandor & Michalakis, 2001), who demonstrated that even in very stable habitats, all organisms will be under selective pressure to disperse some of their progeny. Imagine a population in which the majority of organisms have a stay-at-home, nondispersive genotype O, but in which a rare mutant genotype, X, keeps some offspring at home but commits others to dispersal. The disperser X will suffer no competition in its own patch from O-type individuals but will compete against O-type individuals in their home patches. Disperser X will direct much of its competitive effects at non-kin (with genotype O), while O directs all of its competition at kin (also with genotype O). X will therefore increase in frequency in the population. On the other hand, if the majority of the population are type X, whilst O is the rare mutant, O will still do worse than X, since O can never displace any of the Xs from their patches but has itself to contend with several or many dispersers in its own patch. Dispersal is therefore said to be an evolutionarily stable strategy (ESS) (Maynard Smith, 1972; Parker, 1984). A population of nondispersers will evolve towards the ubiquitous possession of a dispersive tendency; but a population of dispersers will be under no selective pressure to lose that tendency. Hence, the avoidance of both inbreeding and kin competition seem likely to give rise to higher emigration rates at higher densities, when these forces are most intense.

There is indeed evidence for kin competition playing a role in driving offspring away from their natal habitat (Lambin et al., 2001), but much of it is indirect. For example, in the California mouse, Peromyscus californicus, mean dispersal distance increased with increasing litter size in males and, in females, with increasing numbers of sisters in the litter (Ribble, 1992). The more kin a young individual was surrounded by, the further it dispersed.

Lambin et al. (2001) concluded in their review, though, that whereas there is plentiful evidence for density-dependent emigration (see Section 6.3.3), there is little evidence for density-dependent ‘effective’ dispersal (emigration, transfer and immigration), in part at least because immigration (and perhaps transfer) may be inhibited at high densities. For example, in a study of kangaroo rats, Dipodomys spectabilis, over several years during which density varied, dispersal was monitored first after juveniles had become independent of their parents, but then again after they had survived to breed themselves. The kangaroo rats occupy complex burrow systems containing food reserves, and these remain more or less constant in number: high densities therefore mean a saturated environment and more intense competition (Jones et al., 1988). At the time of juvenile independence, density had no effect on dispersal (i.e. on emigration); but by first breeding, dispersal rates (i.e. effective dispersal rates) were lower at higher densities (inverse density dependence) (Figure 6.10). In males, this was mainly because they moved less between juvenile independence and breeding. In females, it occurred mainly because their survival rate in new patches was lower at high densities (Jones, 1988).

6.6.3 Philopatry

Effective dispersal is not straightforwardly density dependent at least in part because there are also selective forces in favor of not dispersing, but instead showing so-called philopatry or ‘home-loving’ behavior (Lambin et al., 2001). This can come about because there are advantages of inhabiting a familiar environment; or individuals may cooperate with (or at least be prepared to tolerate) related individuals in the natal habitat that share a high proportion of their genes; or individuals that do disperse may be confronted with a ‘social fence’ of aggression or intolerance from groups of unrelated individuals (Hestbeck, 1982). These forces, too, may become more intense as the environment becomes more saturated. Thus, for example, Lambin and Krebs (1993) found in Townsend’s voles, Microtus townsendii, in Canada, that the nests or centers of activities of females that were first degree relatives (mother–daughters, littermate sisters) were closer than those that were second degree relatives (nonlittermate sisters, aunt–nieces), which were closer than those that were more distantly related, which in turn were closer than those not related at all.

![Figure 6.10](image-url)

**Figure 6.10** Inverse density-dependent effective dispersal in the kangaroo rat, *Dipodomys spectabilis*: (a) males, (b) females. Natal dispersal distances were greater at low than at high densities. (After Jones, 1988.)
And in a study of Belding’s ground squirrels, *Spermophilus beldingi*, even when females dispersed, they tended to settle near their sisters (Nunes et al., 1997). Moreover, there are examples of fitness being higher when close kin are nearby. For instance, Lambin and Yoccoz (1998) manipulated the relatedness of groups of breeding females of Townsend’s vole, mimicking either a situation where the population had experienced philopatric recruitment followed by high survival (‘high kinship’), or where the population had experienced either low philopatric recruitment or high mortality of recruits (‘low kinship’). Survival of pups, especially early in their life, was significantly higher in the high kinship than in the low kinship treatment.

Overall, then, the relationship between dispersal and density will depend, just like all other adaptations, on evolved compromises to conflicting forces, and also on which aspect of dispersal (emigration, effective dispersal, etc.) is the focus of attention. It is no surprise either that, as we shall see below, the balance of advantage works out differently for different groups: males and females, old and young, and so on. Such variation also argues against broad generalizations suggesting that dispersal is ‘typically’ at presaturation densities (i.e. before resource limitation is intense) or for that matter at saturation densities (Lidicker, 1975).

### 6.7 Variation in dispersal within populations

#### 6.7.1 Dispersal polymorphism

One source of variability in dispersal within populations is a somatic polymorphism amongst the progeny of a single parent. This is typically associated with habitats that are variable or unpredictable. A classic example is the desert annual plant *Gymnarrhena micrantha*. This bears very few (one to three) large seeds (achenes) in flowers that remain unopened below the soil surface, and these seeds germinate in the original site of the parent. The root system of the seedling may even grow down through the dead parent’s root channel. But the same plants also produce above-ground, smaller seeds with a feathery pappus, and these are wind dispersed. In very dry years only the undispersed underground seeds are produced, but in wetter years the plants grow vigorously and produce a large number of seeds above ground, which are released to the hazards of dispersal (Koller & Roth, 1964).

There are very many examples of such seed dimorphism amongst the flowering plants. Both the dispersed and the ‘stay at home’ seeds will, in their turn, produce both dispersed and ‘stay at home’ progeny. Moreover, the ‘stay at home’ seed is often produced from self-pollinated flowers below ground or from unopened flowers, whereas the seeds that are dispersed are more often the product of cross-fertilization. Hence, the tendency to disperse is coupled with the possession of new, recombinant (‘experimental’) genotypes, whereas the ‘stay at home’ progeny are more likely to be the product of self-fertilization.

A dimorphism of dispersers and nondispersers is also a common phenomenon amongst aphids (winged and wingless progeny). As this differentiation occurs during the phase of population growth when reproduction is parthenogenetic, the winged and wingless forms are genetically identical. The winged morphs are clearly more capable of dispersing to new habitats, but they also often have longer development times, lower fecundity, shorter lifespans and hence a reduced intrinsic rate of increase (Dixon, 1998). It is perhaps no surprise, therefore, that aphids may modify the proportions of winged and wingless morphs in immediate response to the environments in which they find themselves. The pea aphid, *Acyrthosiphon pisum*, for example, produces more winged morphs in the presence of predators (Figure 6.11), presumably as an escape response from an adverse environment.

![Figure 6.11](image-url) The mean proportion (± SE) of winged morphs of the pea aphid, *Acyrthosiphon pisum*, produced after two separate periods of exposure to each of two predators: (a) hoverfly larvae and (b) lacewing larvae. Dark bars, predator treatment; light bars, control. (After Kunert & Weisser, 2003.)
6.7.2 Sex-related differences

Males and females often differ in their liability to disperse. Differences are especially strong in some insects, where it is the male that is usually the more active disperser. For example, in the winter moth (*Operophtera brumata*), the female is wingless whilst the male is free-flying. In a seminal paper, Greenwood (1980) contrasted the sex-biased dispersal of birds and mammals. Amongst birds it is usually the females that are the main dispersers, but amongst mammals it is usually the males. Evolutionary explanations for a sex bias have emphasized on the one hand the advantages of a sex bias in its own right as a means of minimizing inbreeding, but also that details of the mating system may generate asymmetries in the costs and benefits of dispersal and philopatry in the two sexes (Lambin et al., 2001). Thus, in birds, competition for territories is typically most intense amongst males. They, therefore, have most to gain from philopatry in terms of being familiar with their natal habitat, whereas the dispersing (and often monogamous) females may gain from exercising a choice of mate amongst the males. In mammals, the (often polygamous) males may compete more often for mates than for territories, and they therefore have most to gain by dispersing to areas with the largest number of defensible females.

6.7.3 Age-related differences

Much dispersal is natal dispersal, i.e. dispersal by juveniles before they reproduce for the first time. In many taxa this is constitutional: we have already noted that seed dispersal in plants is, by its nature, natal dispersal. Likewise, many marine invertebrates have a sessile adult (reproductive) stage and rely on their larvae (obviously pre-reproductive) for dispersal. On the other hand, most insects have a sessile larval stage and rely on the reproductive adults for dispersal. Here, for iteroparous species, dispersal is most often something that occurs throughout the adult life, before and after the first breeding episode; but for semelparous species, dispersal once again is almost inevitably natal.

Birds and mammals, once they have fledged or been weaned and are independent of their mothers, also have the potential to disperse throughout the rest of their lives. None the less, most dispersal here, too, is natal (Wolff, 1997). Indeed, age-biases and sex-biases in dispersal, and the forces of inbreeding-avoidance, competition-avoidance and philopatry, are all tied intimately together in the patterns of dispersal observed in mammals. Thus, for example, in an experiment with gray-tailed voles, *Microtus canicaudus*, 87% of juvenile males and 34% of juvenile females dispersed within 4 weeks of initial capture at low densities, but only 16% and 12%, respectively, dispersed at high densities (Wolff et al., 1997). There was massive juvenile dispersal, which was particularly pronounced in the males; and the inverse density dependence, and especially the very high rates at low densities, argue in favor of inbreeding-avoidance as a major force shaping the pattern.

6.8 The demographic significance of dispersal

The ecological fact of life identified in Section 4.1 emphasized that dispersal can have a potentially profound effect on the dynamics of populations. In practice, however, many studies have paid little attention to dispersal. The reason often given is that emigration and immigration are approximately equal, and they therefore cancel one another out. One suspects, though, that the real reason is that dispersal is usually extremely difficult to quantify.

The nature of the role of dispersal in population dynamics depends on how we think of populations. The simplest view sees a population as a collection of individuals distributed more or less continuously over a stretch of more or less suitable habitat, such that the population is a single, undivided entity. Dispersal is then a process contributing to either the increase (immigration) or decrease (emigration) in the population. Many populations, however, are in fact *metapopulations*; that is, collections of *subpopulations*.

We noted in Section 6.3.1 the ubiquity of patchiness in ecology and the importance of dispersal in linking patches to one another. A subpopulation, then, occupies a habitable patch in the landscape, and it corresponds, in isolation, to the simple view of a population described above. But the dynamics of the metapopulation as a whole is determined in large part by the rate of extinction of individual subpopulations, and the rate of colonization – by dispersal – of habitable but uninhabited patches. Note, however, that just because a species occupies more than one habitable site, each of which supports a population, this does not mean that those populations comprise a metapopulation. As we shall discuss more fully below, ‘classic’ metapopulation status is conferred only when extinction and recolonization play a major role in the overall dynamics.

6.8.1 Modeling dispersal: the distribution of patches

The ways in which dispersal intervenes in the dynamics of populations can be envisaged, or indeed modeled mathematically, in three different ways (see Kareiva, 1990; Keeling, 1999). The first is an ‘island’ or ‘spatially implicit’ approach (Hanski & Simberloff, 1997; Hanski, 1999). Here, the key feature is that a proportion of the individuals leave their home patches and enter a pool of dispersers and are then redistributed amongst patches, usually at random. Thus, these models do not place patches at any specific spatial location. All patches may lose or gain individuals through dispersal, but all are, in a sense, equally distant from all other...
patches. Many metapopulation models, including the earliest (Levins’ model, see below), come into this category, and despite their simplicity (real patches do have a location in space) they have provided important insights, in part because their simplicity makes them easier to analyze.

In contrast, spatially explicit models acknowledge that the distances between patches vary, so do therefore the chances of them exchanging individuals through dispersal. The earliest such models, developed in population genetics, were linear ‘stepping stones’, where dispersal occurred only between adjacent patches in the line (Kimura & Weiss, 1964). More recently, spatially explicit approaches have often involved ‘lattice’ models in which patches are arranged on a (usually) square grid, and patches exchange dispersing individuals with ‘neighboring’ patches – perhaps the four with which they share a side, or the eight with which they make any contact at all, including the diagonals (Keeling, 1999). Of course, despite being spatially explicit, such models are still caricatures of patch arrangements in the real world. They are none the less useful in highlighting new dynamic patterns that appear as soon as space is incorporated explicitly: not only spatial patterns (see, for example, Section 10.5.6), but also altered temporal dynamics, including, for example, the increased probability of extinction of whole spatially explicit metapopulations as habitat is destroyed (Figure 6.12). Further spatially explicit models are also spatially ‘realistic’ (see Hanski, 1999) in that they include information about the actual geometry of fragmented landscapes. One of these, the ‘incidence function model’ (Hanski, 1994b), is utilized below (Section 6.9.4).

Finally, the third approach treats space not as patchy at all but as continuous and homogeneous, and usually models dispersal as part of a reaction–diffusion system, where the dynamics at any given location in space are captured by the ‘reaction’, and dispersal is added as separate ‘diffusion’ terms. The approach has been more useful in other areas of biology (e.g. developmental biology) than it has in ecology. None the less, the mathematical understanding of such systems is strong, and they are particularly good at demonstrating how spatial variation (i.e. patchiness) can be generated, internally, within an intrinsically homogeneous system (Kareiva, 1990; Keeling, 1999).

### 6.8.2 Dispersal and the demography of single populations

The studies that have looked carefully at dispersal have tended to bear out its importance. In a long-term and intensive investigation of a population of great tits, *Parus major*, near Oxford, UK, it was observed that 57% of breeding birds were immigrants rather than born in the population (Greenwood *et al.*, 1978). In a population of the Colorado potato beetle, *Leptinotarsa decemlineata*, in Canada, the average emigration rate of newly emerged adults was 97% (Harcourt, 1971). This makes the rapid spread of the beetle in Europe in the middle of the last century easy to understand (Figure 6.13).

A profound effect of dispersal on the dynamics of a population was seen in a study of *Cakile edentula*, a summer annual plant growing on the sand dunes of Martinique Bay, Nova Scotia. The population was concentrated in the middle of the dunes, and declined towards both the sea and the land. Only in the area towards the sea, however, was seed production high enough and mortality sufficiently low for the population to maintain itself year after year. At the middle and landward sites, mortality exceeded seed production. Hence, one might have expected the population...
to become extinct (Figure 6.14). But the distribution of *Cakile* did not change over time. Instead, large numbers of seeds from the seaward zone dispersed to the middle and landward zones. Indeed, more seeds were dispersed into and germinated in these two zones than were produced by the residents. The distribution and abundance of *Cakile* were directly due to the dispersal of seeds in the wind and the waves.

Probably the most fundamental consequence of dispersal for the dynamics of single populations, though, is the regulatory effect of density-dependent emigration (see Section 6.3.3). Locally, all that was said in Chapter 5 regarding density-dependent mortality applies equally to density-dependent emigration. Globally, of course, the consequences of the two may be quite different. Those that die are lost forever and from everywhere. With emigration, one population’s loss may be another’s gain.

6.8.3 Invasion dynamics

In almost every aspect of life, there is a danger in imagining that what is usual and ‘normal’ is in fact universal, and that what is unusual or eccentric can safely be dismissed or ignored. Every statistical distribution has a tail, however, and those that occupy the tail are as real as the conformists that outnumber them. So it is with dispersal. For many purposes, it is reasonable to characterize dispersal rates and distances in terms of what is typical. But especially when the focus is on the spread of a species into a habitat that it has previously not been occupied, those propagules dispersing furthest may be of the greatest importance. Neubert and Caswell (2000), for example, analyzed the rate of spread of two species of plants, *Calathea ovandensis* and *Dipsacus sylvestris*. In both cases they found that the rate of spread was strongly dependent on the maximum dispersal distance, whereas variations in the pattern of dispersal at lesser distances had little effect.

This dependence of invasion on rare long-distance dispersers means, in turn, that the probability of a species invading a new habitat may have far more to do with the proximity of a source population (and hence the opportunity to invade) than it does on the performance of the species once an initial bridgehead has been established. For instance, the invasion of 116 patches of lowland heath vegetation in southern England by scrub and tree species was studied for the period from 1978 to 1987 (Figure 6.15) and also from 1987 to 1996 (Nolan *et al.*, 1998; Bullock *et al.*, 2002). There were four types of heath – dry, humid, wet and mire – and with
Figure 6.14  Diagrammatic representation of variations in mortality and seed production of Cakile edentula in three areas along an environmental gradient from open sand beach (seaward) to densely vegetated dunes (landward). In contrast to other areas, seed production was prolific at the seaward site. Births, however, declined with plant density, and where births and deaths were equal, an equilibrium population density can be envisaged, $N^*$. In the middle and landward sites, deaths always exceeded births resulting from local seeds, but populations persisted there because of the landward drift of the majority of seed produced by plants on the beach (seaward site). Thus, the sum of local births plus immigrating seeds can balance mortality in the middle and landward sites, resulting in equilibria at appropriate densities. (After Keddy, 1982; Watkinson, 1984.)

Figure 6.15  The invasion (i.e. increase in abundance) of most of the 116 patches of lowland heath in Dorset, UK, by scrub and tree species between 1978 and 1987. Coastland is to the south and the county boundary to the east. (After Bullock et al., 2002.)
two periods, eight data sets on which an analysis could be carried out. For six of these, a significant proportion of the variation in the loss of heath to invading species could be accounted for. The most important explanatory variables were those describing the abundance of scrub and tree species in the vegetation bordering the heath patches. Invasions, and thus the subsequent dynamics of patches, were being driven by initiating acts of dispersal.

6.9 Dispersal and the demography of metapopulations

6.9.1 The development of metapopulation theory: uninhabited habitable patches

Recognition that many populations are in fact metapopulations was firmly established around 1970, but there was a delay of around 20 years before that recognition was translated into action and an increasing number of studies placed metapopulation dynamics prominently on the ecological stage. Now, the danger is not so much one of neglect, but that all populations are thought of as metapopulations, simply because the world is patchy.

Central to the concept of a metapopulation is the idea, emphasized by Andrewartha and Birch back in 1954, that habitable patches might be uninhabited simply because individuals have failed to disperse into them. To establish that this is so, we need to be able to identify habitable sites that are not inhabited. Only very rarely has this been attempted. One method involves identifying characteristics of habitat patches to which a species is restricted and then determining the distribution and abundance of similar patches in which the species might be expected to occur. The water vole (*Arvicola terrestris*) lives in river banks, and in a survey of 39 sections of river bank in North Yorkshire, UK, 10 contained breeding colonies of voles (core sites), 15 were visited by voles but they did not breed there (peripheral sites) and 14 were apparently never used or visited. A ‘principle component’ analysis was used to characterize the core sites, and on the basis of these characteristics a further 12 unoccupied or peripheral sites were identified that should have been suitable for breeding voles (i.e. habitable sites). Apparently, about 30% of habitable sites were uninhabited by voles because they were too isolated to be colonized or in some cases suffered high levels of predation by mink (Lawton & Woodroffe, 1991). Habitable patches can also be identified for a number of rare butterfly species because the larvae feed on only one or a few patchily distributed plant species. Thomas et al. (1992) found that the patches that remained uninhabited were small and isolated from the sources of dispersal: the butterfly *Phlebejus argus* was able to colonize virtually all habitable sites less than 1 km from existing populations. Indeed, the habitability of some of the isolated (previously uninhabited) sites was established when the butterfly was successfully introduced (Thomas & Harrison, 1992). This is the crucial test of whether a site is really habitable or not.

6.9.2 The development of metapopulation theory: islands and metapopulations

The classic book, *The Theory of Island Biogeography* by MacArthur and Wilson (1967), was an important catalyst in radically changing ecological theory in general. The authors developed their ideas in the context of the dynamics of the animals and plants on real (maritime) islands, which they interpreted as reflecting a balance between the opposing forces of extinctions and colonizations. They emphasized that some species (or local populations) spend most of their time either recovering from past crashes or in phases of invasion of new territories (islands), while others spend much of their time at or around their carrying capacity. These two ends of a continuum are the *r* and *K*-species of Section 4.12. At one extreme (*r*-species), individuals are good colonizers and have characteristics favoring rapid population growth in an empty habitat. At the other end of the continuum (*K*-species) individuals are not such good colonizers but have characteristics favoring long-term persistence in a crowded environment. *K*-species therefore have relatively low rates of both colonization and extinction, whereas *r*-species have relatively high rates. These ideas are developed further in the discussion of island biogeography in Chapter 21.

At about the same time as MacArthur and Wilson’s book was published, a simple model of ‘metapopulation’ dynamics was proposed by Levins (1969, 1970). Like MacArthur and Wilson, he sought to incorporate into ecological thinking the essential patchiness of the world around us. MacArthur and Wilson were more concerned with whole communities of species, and envisaged a ‘mainland’ that could provide a regular source of colonists for the islands. Levins focused on populations of a single species and awarded none of his patches special mainland status. Levins introduced the variable *p(t)*, the fraction of habitat patches occupied at time *t*, reflecting an acceptance that not all habitable patches are always inhabited.

The rate of change in the fraction of occupied habitat (patches, *p*) is given in Levins’ model as:

\[
\frac{dp}{dt} = mp(1-p) - \mu p, \tag{6.1}
\]

in which \(\mu\) is the rate of local extinction of patches and \(m\) is the rate of recolonization of empty patches. That is, the rate of recolonizations increases both with the fraction of empty patches prone to recolonization \((1-p)\) and with the fraction of occupied patches able to provide colonizers, \(p\), whereas the rate of extinctions increases simply with the fraction of patches prone to extinction, \(p\). Rewriting this equation, Hanski (1994a) showed...
that it is structurally identical to the logistic equation (see Section 5.9):

\[
\frac{dp}{dt} = (m - \mu) \ p \ (1 - p/(1 - (m/\mu))).
\]

(6.2)

Hence, as long as the intrinsic rate of recolonization exceeds the intrinsic rate of extinction \((m - \mu > 0)\), the total metapopulation will reach a stable equilibrium, with a fraction, \(1 - (\mu/m)\), of the patches occupied.

The most fundamental message from taking a metapopulation perspective, then, which emerges from even the simplest models, is that a metapopulation can persist, stably, as a result of the balance between random extinctions and recolonizations even though none of the local populations are stable in their own right. An example of this is shown in Figure 6.16, where within a persistent, highly fragmented metapopulation of the Glanville fritillary butterfly \((Melitaea cinxia)\) in Finland, even the largest local populations had a high probability of declining to extinction within 2 years. To restate the message another way: if we wish to understand the long-term persistence of a population, or indeed that population’s dynamics, then we may need to look beyond the local rates of birth and death (and what determines them), or even the local rates of immigration and emigration. If the population as a whole functions as a metapopulation, then the rates of subpopulation extinction and colonization may be of at least comparable importance.

**Figure 6.16** Comparison of the local population sizes in June 1991 (adults) and August 1993 (larvae) of the Glanville fritillary butterfly \((Melitaea cinxia)\) on Åland island in Finland. Multiple data points are indicated by numbers. Many 1991 populations, including many of the largest, had become extinct by 1993. (After Hanski et al., 1995.)

6.9.3 When is a population a metapopulation?

Two necessary features of a metapopulation have already been established here: that individual subpopulations have a realistic chance of experiencing both extinction and recolonization. To this we can add a third, which has been implicit in the discussion thus far. The dynamics of the various subpopulations should be largely independent, i.e. not synchronous. There would, after all, be little hope of stability if when one subpopulation went extinct they all did. Rather, asynchrony guarantees that as one goes extinct (or even declines), there are likely to be others that are thriving and generating dispersers, promoting the ‘rescue effect’ (Brown & Kodric-Brown, 1977) of the former by the latter.

Some metapopulations may conform to the ‘classic’ concept, in which all the subpopulations have a realistic (and roughly equal) chance of extinction, but in other cases there may be significant variation in either the size or quality of individual patches. Thus, patches may be divided into ‘sources’ (donor patches) and ‘sinks’ (receiver patches) (Pulliam, 1988). In source patches at equilibrium, the number of births exceeds the number of deaths, whereas in sink patches the reverse is true. Hence, source populations support one or more sink populations within a metapopulation. The persistence of the metapopulation depends not only on the overall balance between extinction and recolonization, as in the simple model, but also on the balance between sources and sinks.

In practice, of course, there is likely to be a continuum of types of metapopulation: from collections of nearly identical local populations, all equally prone to extinction, to metapopulations in which there is great inequality between local populations, some of which are effectively stable in their own right. This contrast is illustrated in Figure 6.17 for the silver-studded blue butterfly \((Plejebus argus)\) in North Wales.

Just because a population is patchily distributed, however, this does not necessarily make it a metapopulation (Harrison & Taylor, 1997; Bullock et al., 2002). First, a population may be patchily distributed, but dispersal between the patches may be so great that the dynamics of the individual patches are no longer independent: a single population, albeit occupying a heterogeneous habitat. Alternatively, patches may be so isolated from one another that dispersal between them is negligible: a series of effectively separate populations.

Finally, and perhaps most commonly, all patches may simply have a negligible chance of extinction, at least on observable timescales. This means that their dynamics may be influenced by birth, death, immigration and emigration – but not to any significant degree by extinction or recolonization. This last category comes closest to being a true metapopulation, and there can be little doubt that the title has been given to many patchy populations fitting this description. Of course, there can be a danger in being overprotective of the purity of definitions.
What harm can there be if, as interest in the metapopulation concept grows, the term itself is extended to a wider variety of ecological scenarios? Perhaps none – and the spread of the term’s usage to populations originally beyond its reach may, in any case, be unstoppable. But a word, like any other signal, is only effective if the receiver understands what the sender intends. At the very least, care should be taken by users of the term to confirm whether the extinction and recolonization of patches has been established.

The problem of identifying metapopulations is especially apparent for plants (Husband & Barrett, 1996; Bullock et al., 2002). There is no doubt that many plants inhabit patchy environments, and apparent extinctions of local populations may be common. This is shown in Figure 6.18 for the annual aquatic plant *Eichhornia paniculata*, living in temporary ponds and ditches in arid regions in northeast Brazil. However, the applicability of the idea of recolonization following a genuine extinction is questionable in any plant species that has a buried seed bank. In *E. paniculata*, for instance, the heavy seeds almost always drop in the immediate vicinity of the parent rather than being dispersed to other patches. ‘Extinctions’ are typically the result of the catastrophic loss of habitat (note in Figure 6.18 that the chance of extinction has effectively nothing to do with the previous population size) and ‘recolonizations’ are almost always simply the result of the germination of seeds following habitat restoration. Recolonization by dispersal, a prerequisite for a true metapopulation, is extremely rare.

Moreover, as Bullock et al. (2002) point out, of the plant studies that have documented patch extinctions and colonizations,
the vast majority have been in recently emerged patches (the early stages of succession, see Chapter 16). Extinctions mostly occur when the vegetation in a patch develops to a state where it is no longer suitable for the plant species in question, and that patch is therefore also not suitable for recolonization by the same species. This is ‘habitat tracking’ (Harrison & Taylor, 1997) rather than the repeated extinction and recolonization of the same habitat that is central to the concept of a metapopulation.

6.9.4 Metapopulation dynamics

Levins’ simple model does not take into account the variation in size of patches, their spatial locations, nor the dynamics of populations within individual patches. Not surprisingly, models that do take all these highly relevant variables into account become mathematically complex (Hanski, 1999). Nevertheless, the nature and consequences of some of these modifications can be understood without going into the details of the mathematics.

For example, imagine that the habitat patches occupied by a metapopulation vary in size and that large patches support larger local populations. This allows persistence of the metapopulation, with lower rates of colonization than would otherwise be the case, as a result of the lowered rates of extinction on the larger patches (Hanski & Gyllenberg, 1993). Indeed, the greater the variation in patch size, the more likely it is that the metapopulation will persist, other things being equal. Variations in the size of local populations may, alternatively, be the result of variations in patch quality rather than patch size: the consequences would be broadly the same.

The probability of extinction of local populations typically declines as local population size increases (Hanski, 1991). Moreover, as the fraction of patches occupied by the metapopulation, $p$, increases, there should on average be more migrants, more immigration into patches, and hence larger local populations (confirmed, for example, for the Glanville fritillary – Hanski et al., 1995). Thus, the extinction rate, $\mu$, should arguably not be constant as it is in the simple model, but should decline as $p$ increases. Models incorporating this effect (Hanski, 1991; Hanski & Gyllenberg, 1993) often give rise to an intermediate unstable threshold value of $p$. Above the threshold, the sizes of local populations are sufficiently large, and their rate of extinction sufficiently low, for the metapopulation to persist at a relatively high fraction of patches, as in the simple model. Below the threshold, however, the average size of local populations is too low and their rate of extinction hence too high. The metapopulation declines either to an alternative stable equilibrium at $p = 0$ (extinction of the whole metapopulation) or to one in which $p$ is low, where essentially only the most favorable patches are occupied.

Different metapopulations of the same species might therefore be expected to occupy either a high or a low fraction of their habitable patches (the alternative stable equilibria) but not an intermediate fraction (close to the threshold). Such a bimodal distribution is indeed apparent for the Glanville fritillary in Finland (Figure 6.19). In addition, these alternative equilibria have potentially profound implications for conservation (see Chapter 15), especially when the lower equilibrium occurs at $p = 0$, suggesting that the threat of extinction for any metapopulation may increase or decline quite suddenly as the fraction of habitable patches occupied moves below or above some threshold value.

One study drawing many of the preceding threads together examined the dynamics of a presumed metapopulation of a small mammal, the American pika Ochotona princeps, in California (Moilanen et al., 1998). (The qualifier ‘presumed’ is necessary because dispersal between habitat patches was itself presumed rather than actually observed (see Clinchy et al., 2002).) The overall metapopulation could itself be divided into northern, middle and southern networks, and the patch occupancy in each was determined on four occasions between 1972 and 1991 (Figure 6.20a). These purely spatial data were used alongside more general information on pika biology, to provide parameter values for Hanski’s (1994b) incidence function model (see Section 6.8.1). This was then used to simulate the overall dynamics of each of the networks, with a realistic degree of stochastic variation incorporated, starting from the observed situation in 1972 and either treating the entire metapopulation as a single entity (Figure 6.20b) or simulating each of the networks in isolation (Figure 6.20c).
The data themselves (Figure 6.20a) show that the northern network maintained a high occupancy throughout the study period, the middle network maintained a more variable and much lower occupancy, while the southern network suffered a steady and substantial decline. The output from the incidence function model (Figure 6.20b) was very encouraging in mirroring accurately these patterns in temporal dynamics despite being based only on spatial data. In particular, the southern network was predicted to collapse periodically to overall extinction but to be rescued by the middle network acting, despite its low occupancy, as a stepping stone from the much more buoyant northern network. This interpretation is supported by the results when the three networks are simulated in isolation (Figure 6.20c). The northern network remains at a stable high occupancy; but the middle network, starved of migrants from the north, rapidly and predictably crashes; and the southern network, while not so unstable, eventually suffers the same fate. On this view, then, within the metapopulation as a whole, the northern network is a source and the middle and southern networks are sinks. Thus, there is no need to invoke any environmental change to explain the decline in the southern network; such declines are predicted even in an unchanging environment.

Even more fundamentally, these results illustrate how whole metapopulations can be stable when their individual subpopulations are not. Moreover, the comparison of the northern and middle networks, both stable but at very different occupancies, shows how occupancy may depend on the size of the pool of dispersers, which itself may depend on the size and number of the subpopulations.

Finally, these simulations direct us to a theme that recurs throughout this book. Simple models (and one’s own first thoughts) often focus on equilibria attained in the long term. But in practice such equilibria may rarely be reached. In the present case, stable equilibria can readily be generated in simple metapopulation models, but the observable dynamics of a species may often have more to do with the ‘transient’ behavior of its metapopulations, far from equilibrium. To take another example, the silver-spotted skipper butterfly (*Hesperia comma*) declined steadily in Great Britain from a widespread distribution over most calcareous hills in 1900, to 46 or fewer refuge localities (local populations) in 10 regions by the early 1960s (Thomas & Jones, 1993). The probable reasons were changes in land use – increased ploughing of unimproved grasslands and reduced stocking with...
domestic grazing animals – and the virtual elimination of rabbits by myxomatosis with its consequent profound vegetational changes. Throughout this nonequilibrium period, rates of local extinction generally exceeded those of recolonization. In the 1970s and 1980s, however, the reintroduction of livestock and the recovery of the rabbits led to increased grazing and the number of suitable habitats increased again. Recolonization exceeded local extinction – but the spread of the skipper remained slow, especially into localities isolated from the 1960s refugia. Even in southeast England, where the density of refugia was greatest, it is predicted that the abundance of the butterfly will increase only slowly – and remain far from equilibrium – for at least 100 years.

Summary

We distinguish between dispersal and migration, and within dispersal between emigration, transfer and immigration.

Various categories of active and passive dispersal are described, including especially passive dispersal in the seed rain and the guerrilla and phalanx strategies of clonal dispersers.

Random, regular and aggregated distributions are explained, and the importance of scale and patchiness in the perception of such distributions is emphasized, especially in the context of environmental ‘grain’. Forces favoring and diluting aggregations are elaborated, including the theory of the selfish herd and density-dependent dispersal.

We describe some of the main patterns of migration at a range of scales – tidal, diurnal, seasonal and intercontinental – including those that recur repeatedly and those that occur just once.

We examine dormancy as migration in time in both animals (especially diapause) and plants. The importance of photoperiod in the timing of dormancy is emphasized.

The relationship between dispersal and density is examined in detail. The roles of in- and outbreeding in driving density dependences are explained, including especially the importance of avoiding kin competition on the one hand and the attractions of philopatry on the other.

We describe a variety of types of variation in dispersal within populations: polymorphisms and sex- and age-related differences.

We turn to the demographic significance of dispersal and introduce the concept of the metapopulation composed of a number of subpopulations. Dispersal can be incorporated into the dynamics of populations, and modeled, in three different ways: (i) an ‘island’ or ‘spatially implicit’ approach; (ii) a spatially explicit approach that acknowledges that the distances between patches vary; and (iii) an approach treating space as continuous and homogeneous.

Probably the most fundamental consequence of dispersal for the dynamics of single populations is the regulatory effect of density-dependent emigration. It is important also, though, to recognize the importance of rare long-distance dispersers in invasion dynamics.

Metapopulation theory developed from the earlier concept of the uninhabited habitable patch. Its origin as a concept in its own right was the Levins’ model, which established the most fundamental message: that a metapopulation can persist, stably, as a result of the balance between random extinctions and recolonizations, even though no subpopulations are stable in their own right.

Not all patchily distributed populations are metapopulations, so we address the question ‘When is a population a metapopulation?’, which may be particularly problematic with plant populations.

Finally, we explore the dynamics of metapopulations, emphasizing especially the likely importance of alternative stable equilibria.