A

ADAPTIVE BEHAVIOR AND VIGILANCE

PETER A. BEDNEKOFF
Eastern Michigan University, Ypsilanti

Adaptive behavior is behavior that raises an animal's fitness in its biotic and abiotic environment. The study of adaptive behavior is mainly the study of how behavior changes with environmental changes. Fitness is the relative contribution of an organism to genes in future generations. Because understanding fitness across the entire life of an organism is a daunting task (and tracking it across several generations even more so), researchers have commonly assumed a particular relationship between behavior in the short term and fitness. This has involved examining many behaviors using many fitness proxies and modeling techniques. This entry explores the simplifying assumptions about the relationship between behavior and fitness while concentrating on a single type of behavior, watching for predators. Though the focus here is on such vigilance behavior, the approach employed may be applied to many other types of behavior.

BASELINE MODELS OF VIGILANCE BEHAVIOR

Anti-predator vigilance involves pauses in other behavior, such as feeding, in order to scan the environment for predators. Researchers commonly assume that that foraging animals lift their heads independently of one another and share information about detected attacks. Shared information about attacks is known as collective detection. The original model of anti-predator vigilance showed that in bigger groups each individual could scan less while maintaining the same probability of an undetected attack. If each individual is vigilant v proportion of the time, then an attack goes undetected when no individual is vigilant, which occurs a fraction $(1 - v)^n$ of the time, where n is the number of individuals in the group. This model did not relate this directly to fitness but determined that animals could scan less in larger groups at the same risk of an undetected attack. Subsequent studies have built upon the assumptions of independent scanning and collective detection of attacks but added dilution of risk. Here, the predator can only effectively attack one individual (or perhaps a small portion of the group), even when all individuals are unaware of the impending attack.

Other researchers have supposed that animals maximize survival while maintaining a required level of food intake. This second criterion can make sense if food increases do not impact future reproduction very much, for example, during the nongrowing season for animals that have small growth during this period. These two criteria for adaptive behavior—maximizing food intake while maintaining a set probability of surviving and minimizing risk of attack and maintaining a set level of food intake—obviously cannot both hold true simultaneously. They can, however, both be approximations of a larger truth. In general, fitness will depend on the value of food and the probability of surviving. In simple models, vigilance increases survival but decreases food intake, and thereby future reproductive output. This entry discusses models of the optimal level of vigilance that build upon the work of Parker and Hammerstein but differ from their models by allowing multiple attacks and assuming future reproduction is a linearly increasing function of foraging intake.

To develop heuristic models of vigilance in groups, assume that foragers feed for some extended time, T, before reproducing. Animals can forage at any rate between 0 and 1, and there is a direct tradeoff with vigilance—food intake is proportional to $1 - v$. An animal's total foraging intake across the period is thus proportional to $(1 - v)T$, and reproductive success at the end of the period is proportional to foraging intake—$V = k(1 - v)T$. In order to reproduce, however, animals must survive this period. If they are attacked at rate α according to a Poisson process, then the probability of survival decreases exponentially with the attack rate, period length, and the mortality per attack—$S = \text{Exp}[-\alpha TM]$. The probability of dying in an attack increases with foraging rate. For a single individual the mortality rate is $M = (1 - v)$. In a group, the mortality rate may depend on the actions of others. This is the topic of the next section.

PERFECT COLLECTIVE DETECTION

When a predator attacks a group, the focal individual dies if it is not vigilant and is not warned by other members of the group. For the effects of others, the classic assumption is perfect collective detection: the focal individual is warned of the attack if any group member is vigilant at the time. Thus, the attack succeeds only if no member of the group is vigilant at the time. If the focal individual is joined by $n - 1$ other individuals, each vigilant $\hat{v}$ of the time, this equals $(1 - v)(1 - \hat{v})^{n-1}$. If the predator must choose among these n unaware prey and does so equally, the risk is $1/n$ for each group member. Thus, the probability that the focal individual dies in an attack is

$$M = \frac{(1 - v)(1 - \hat{v})^{n-1}}{n}.$$

As we would expect, animals are safer when they are more vigilant and when they are in larger groups. Because the fitness of the focal individual depends on the vigilance of others, game theory is needed to find the solution. The evolutionarily stable strategy (ESS) can be found by differentiating the fitness function with respect to v, setting it to zero, and solving for v. This gives v as a function of $\hat{v}$—in this case:

$$v^* = 1 - \frac{n}{\alpha T(1 - \hat{v})^{n-1}}$$

Because an individual can be warned by others, it can rely on their vigilance to some extent. In a group where everyone else is very vigilant, the best response is to be less vigilant. In a group where others are not vigilant, the best response is to be more vigilant. In between, there is a level of vigilance that is the best response to itself. To find this evolutionarily stable strategy, set $\hat{v} = v$ and again find v. In this case the evolutionarily stable level of vigilance is

$$v^* = 1 - \left(\frac{n}{\alpha T}\right)^{\frac{1}{n}}.$$

The evolutionarily stable strategy here is equivalent to the Nash equilibrium from game theory. For comparison, the optimal cooperative strategy, or Pareto equilibrium, can be found by substituting $\hat{v} = v$ in the fitness function, differentiating, setting this equal to zero, and then solving for v. With this first model, the optimal cooperative strategy is

$$v_c^* = 1 - \left(\frac{1}{\alpha T}\right)^{\frac{1}{n}}.$$

These expressions differ only in the numerator of the ratio in the second right-hand term. Since this second right-hand term equals the feeding rate, individuals feed at $n^{1/n}$ times the cooperative rate under the ESS. These results replicate findings that ESS levels of vigilance are lower than cooperative when collective detection is perfect. Analyzing $n^{1/n}$ as n varies shows that the difference between the ESS and cooperative solutions is greatest for $n = 3$ and declines with larger group sizes.

NO COLLECTIVE DETECTION

A null model indicates how anti-predator behavior would be different without collective detection. In this model, individuals that are vigilant at the start of the attack escape, and the predator targets one of the nondetectors. As the number of detectors, i, goes up, the effective group size for dilution decreases by the same number. As before, our focal individual is in danger only when it is not vigilant $(1 - v)$ but now the effects of others must be summed across all possible numbers of detectors:

$$M = (1 - v)\sum_{i=0}^{n-1}\left(\frac{(n-1)!}{(n-1-i)!i!}\right)\frac{\hat{v}^i(1 - \hat{v})^{n-1-i}}{n - i}.$$

The possible number of detectors ranges from zero to $n - 1$—all group members other than our focal individual. Inside the summation, the factorial gives the number of ways of having i detectors, the numerator of the right-hand term gives the probability of any one such combination, and the denominator gives the dilution of risk among the $n-1$ nondetectors (including the focal individual). This sum simplifies such that the overall mortality is

$$M = \frac{(1 - v)(1 - \hat{v}^n)}{n(1 - \hat{v})}.$$

Once again the optimal response depends on the vigilance of others,

$$v^* = 1 - \frac{n(1 - \hat{v})}{\alpha T(1 - \hat{v}^n)},$$

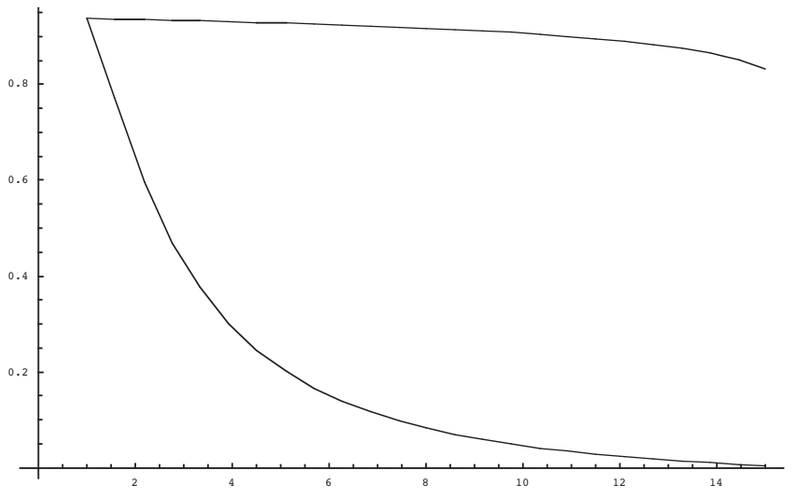


FIGURE 1 The evolutionarily stable level of vigilance declines steeply with group size when collective detection is perfect (lower curve) but very weakly when collective detection is absent (upper curve). Here the expected number of attacks (αT) is 16.

and the ESS occurs in the case when the optimal response is to match the vigilance of others, which yields

$$v^* = \left(1 - \frac{n}{\alpha T}\right)^{\frac{1}{n}}.$$

This expression is negative or undefined when the expected number of attacks is less than the group size, $\alpha T < n$. In this situation, zero vigilance is the best solution.

In this null model, a group size effect is present, but generally weak (Fig. 1). Since a strong effect of group size on vigilance has been observed many times in nature, this result suggests that the observed group size effect on vigilance probably depends upon collective detection to some extent.

The cooperative solution for the null model is difficult to state in general, but for a pair of animals it is

$$v_c^* = \frac{1}{2} + \sqrt{\frac{\alpha T - 4}{2\alpha T}}.$$

When fewer than four attacks occur on average ($\alpha T < 4$), this expression is undefined and zero vigilance ($v = 0$) is the best available vigilance level. Compared to the ESS solution for a pair (Fig. 2), without collective detection the cooperative vigilance level is lower than the ESS vigilance level.

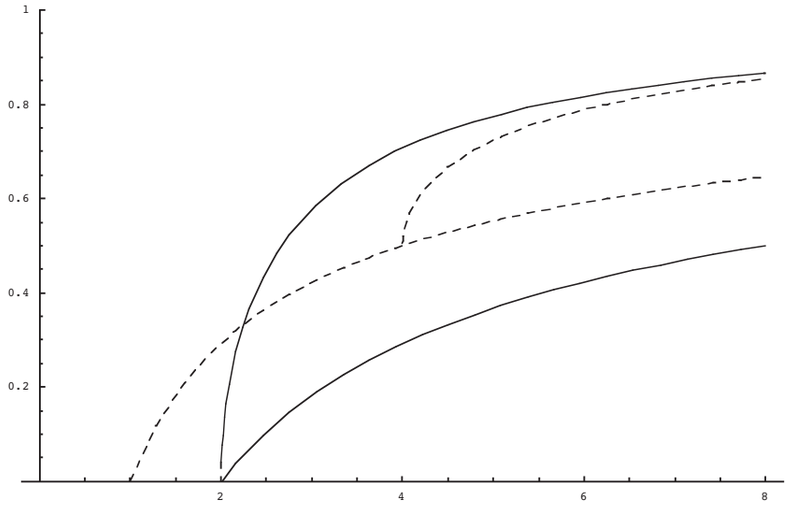


FIGURE 2 Vigilance levels increase with the expected number of attacks whether vigilance is cooperative or not, and whether collective detection is perfect or absent. The upper two curves are when collective detection is absent. Here ESS levels (solid lines) are higher than cooperative levels (dashed lines). The lower two curves assume collective detection is perfect. Here cooperative vigilance levels (dashed lines) are higher than ESS levels (solid lines). All results are for pairs of animals.

The models in this entry differ from those of Parker and Hammerstein first in allowing for multiple attacks rather than a single attack. Their equivalent model produced an ESS of zero vigilance for all group sizes greater than 1. In the models described here, which include perfect collective detection, the ESS level of vigilance falls to zero when the number of expected attacks is no bigger than the group size. Thus, the two models produce similar results under the correct simplifying assumptions, but quite different results are possible when multiple attacks are considered.

While earlier models suggested ESS vigilance should be zero for pairs or larger groups, observations showed nonzero levels of vigilance in groups. This mismatch led to consideration of cooperative models of vigilance. The conclusions of these early models apply only in some cases. With perfect collective detection, cooperative vigilance is generally higher than the ESS level. Without collective detection, however, cooperative vigilance is generally lower. In the model with no collective detection, the cooperative response is zero vigilance, whereas the ESS solution is moderate levels of vigilance when between two and four attacks are expected (Fig. 2) and the ESS level of vigilance is generally higher than the cooperative level anytime two or more attacks are expected.

PARTIAL COLLECTIVE DETECTION

The models in this entry have thus far only examined the extremes of perfect collective detection and no collective detection. Empirical studies indicate that collective detection is real yet slow and potentially faulty. For large groups, a model of imperfect collective detection would need to keep track of the overall contagion of information. This has not been done and is beyond the scope of this entry. For a pair of animals, however, a model only needs to track the flow of information from a primary detector to the other member of the pair. Here, the danger for the focal animal is

$$M = (1 - v)\left(\frac{(1 - \hat{v})}{2} + \hat{v}(1 - i)\right).$$

The focal animal is in danger when it fails to detect an attack $(1 - v)$. If the other animal fails to detect the attack $(1 - \hat{v})$, the two are equally likely to be targeted by the predator. If the other animal detects the attack (with probability $\hat{v}$), the focal animal escapes if it gets information, but is the sole target for attack if it fails to get information $(1 - i)$ from the other individual, where i is the probability that the focal animal receives a warning from the other animal.

Study of vigilance generally includes the separate advantages for collective detection and risk dilution, where risk dilution occurs when the predator chooses among of a group of uninformed prey. This distinction blurs, however, when collective detection is imperfect and takes time. If different members of the group have different amounts of information during the course of the attack, dilution depends strongly on when and how predators target prey for attack. If predators target prey early in an attack, risk is spread equally among the n group members. If predators wait until later in the attack, however, they can choose among the subset of the group that has not yet detected them. Thus, dilution depends on how many other individuals are similarly uninformed of the attack. The formulation given here assumes that the predator chooses a target after primary detection but before secondary detection is complete. Thus, the predator ignores any individual with their head up at the time of the attack, and chooses with equal probability among the other individuals. This gives an advantage to individuals who are vigilant at the start of the attack. Other formulations would be appropriate if predators could target individuals that were the very last to become informed. Such formulations would place even greater advantage on personal detection of attacks.

Working with the mortality expression for imperfect collective, we solve and find that the ESS level of vigilance for a member of a pair is

$$v^* = \left(\frac{1}{2i - 1}\right)\left(\frac{i - \sqrt{4i - 2 + \alpha T(1 - i)^2}}{\sqrt{\alpha T}}\right),$$

while the cooperative level is

$$v_c^* = \left(\frac{1}{4i - 2}\right)\left(\frac{3i - 1 - \sqrt{8i - 4 + \alpha T(1 - i)^2}}{\sqrt{\alpha T}}\right).$$

For small numbers of attacks, these expressions may be negative—so that the internal optimum is zero vigilance. With more attacks, vigilance tends to increase under either the ESS or the cooperative criterion. The effects of information flow are more complicated. With increasing information passing from detector to nondetector, the ESS level of vigilance generally decreases. Cooperative vigilance, however, increases with greater information flow at moderately low attack rates but then decreases at higher attack rates. In all cases, cooperative vigilance is higher than noncooperative vigilance only with $i > 1/2$ (Fig. 3). Other work has shown that $i > 1/2$ is also the criterion for coordinated vigilance to be advantageous. As pointed out by Steve Lima, the widespread observation that vigilance is not coordinated is evidence against

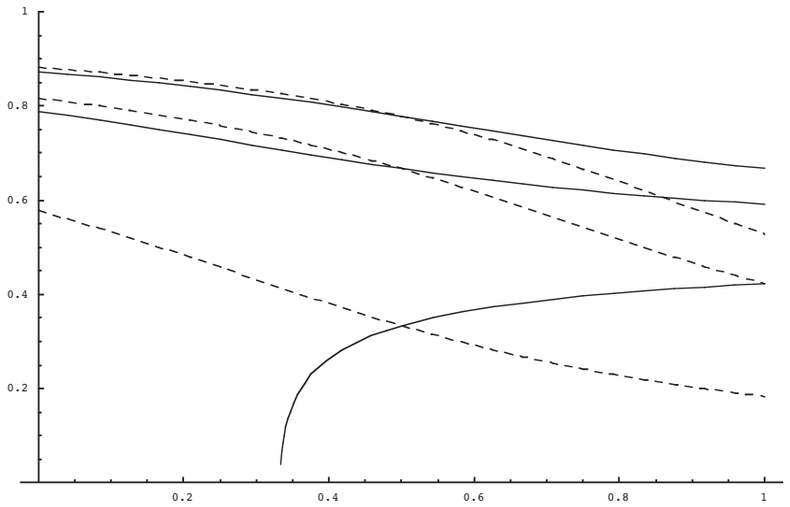


FIGURE 3 ESS and cooperative levels of vigilance as a function of information (i) passing from an animal detecting an attack to one that has not. The three pairs of curves are for 3, 6, and 9 expected attacks. Vigilance levels increase when more attacks are expected. Within each pair of curves, cooperative levels of vigilance (dashed lines) are lower than ESS levels (solid lines) for $i < 1/2$ and higher for $i > 1/2$.

cooperation. The potential fruit of cooperation, the public goods in this case, depend critically on the information that nondetectors get from detectors of attacks. In nature this information may be incomplete or unreliable, so each individual needs to gather information for itself. Individual vigilance is both more valuable and more costly than relying on social information. Thus with multiple attacks and imperfect collective detection, the norm is for each individual to invest in vigilance.

The ESS level of vigilance decreases with greater information passing from detector to nondetector. This is seen by examining the best response to the vigilance of the other group member (Fig. 4). With little or no information, the best reaction approaches the level for a single bird (7/8). When information only has a probability one-half of reaching a nondetector, the reaction does not depend on the vigilance of the other group member. Here any increase in detection with the vigilance of the other animal is exactly offset by decrease in dilution when information fails to get through. With $i > 1/2$, the best response is considerably lower than for a lone animal, and lower when the other animal is more vigilant. Observations show that

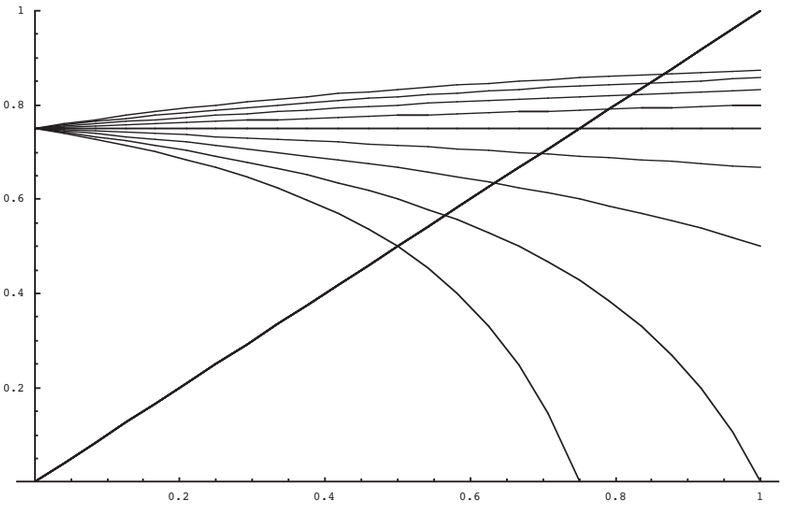


FIGURE 4 Optimal vigilance level in response to vigilance of the other animal in a pair. The curves differ in the probability of information about an attack passing from a detecting individual to a nondetector. The top curve is for no information ($i = 0$) and the bottom one for perfect information ($i = 1$) with i changing by 1/8 for each curve in between. ESS levels of vigilance lie along the diagonal line where the best response equals the vigilance of the other animal. The expected number of attacks (αT) equals 8.

animals are often much less vigilant in pairs than alone, suggesting considerable information flow and collective detection. On the other hand, there is much less information that animals react directly to the vigilance rates of others. Perhaps the danger to one member of a group is little influenced by the vigilance rates of others, either because information flow is moderate ($i \approx 1/2$) or due to some other factor not included in this model.

RISK ALLOCATION

Classic models of vigilance behavior generally considered perfect collective detection and a very limited number of attacks. Some outcomes of the models differ with imperfect collective detection and multiple attacks. Multiple attacks might occur over a substantial period of time, and conditions might change during this period. Consider the situation when group size changes over time. For simplicity, assume that the focal animal spends some portion of its time, p, alone and the rest of the time, q, paired with some other animal. Other assumptions are the same as above. If individuals face various situations with different levels of risk, they may allocate vigilance and feeding behavior across the situations, with v_1 being the vigilance in situation 1 and v_2 the vigilance in situation 2.

In contrast to previous models of risk allocation, the models presented here use a linear relationship between feeding rate and detection behavior. This assumption simplifies the models, and also leads to the solution of zero vigilance in the safer situation. The results are striking in that the safer situation in a pair switches at $i = 1/2$. This model of risk allocation does not yield a simple ratio of feeding rates, because one feeding rate is always the maximum, 1. The model, nonetheless, illustrates that vigilance depends upon the schedule of other situations an animal faces. With $i < 1/2$, the ESS level of vigilance is zero when alone, and $1 - (1/\alpha Tp)$ when in a pair. Thus, allocation has reversed the modest group size effect seen above with little collective detection. With moderate to high collective detection ($i > 1/2$), the vigilance in the pair goes to zero with the vigilance of an animal feeding alone, depending on how often it is in a pair:

$$v_2^* = 0,\ v_1 = \frac{2\alpha T - 1}{2\alpha Tp}$$

The result is remarkable and counterintuitive in that collective detection leads to an evolutionary stable solution where neither individual is vigilant, detection does not occur, and so collective detection never happens. Thus in any attack on a pair, one individual or the other will die. Because this model assumes that the proportion of time in a pair is not affected by vigilance levels, it implicitly assumes that group members that die are replaced before the next attack. The next section explores whether this combination of assumptions leads to the surprising results.

PARTNER LOSS AND MUTUAL DEPENDENCE

Most models of vigilance have assumed a stable group size. Models of risk allocation assume changes in risk factors, such as group size, but that the schedule of risk factor changes is not influenced by their feeding or vigilance rates. Now consider a model where group size changes as a result of mortality. Assume all individuals start in pairs and that collective detection is perfect ($i = 1$). The previous results for risk allocation provide an interesting foundation in that they do not predict vigilance in pairs. Therefore vigilance in pairs is due to protecting an irreplaceable partner—which is one kind of mutual dependence in fitness.

The assumptions for what happens when a predator attacks a pair are exactly as before. If the other animal dies in an attack, however, our focal animal faces all future attacks alone. The solution involves two vigilance rates, one in a pair and one alone. Survival is summed across all routes in which the focal individual survives, each weighted by its probability. To find foraging success, note that if when the other animal dies in the ith of j attacks, the focal animal forages in a pair for on average $i/(j + 1)$ of the total time period and alone the rest of the period:

$$S = e^{-\alpha} + \sum_{i=1}^{\infty} \frac{\alpha^i e^{-\alpha}}{i!} \sum_{j=1}^{i} (v_2 + \hat{v}_2 + v_2\hat{v}_2)^{j-1} \times \left(\frac{(1 - v_2)(1 - \hat{v}_2)}{2}\right)^{j} (1 - v_1)^{i-j} k\left(v_2 \frac{j}{i+1} + v_1 \frac{1-j}{i+1}\right)$$

The most important result of this model is that the ESS result is to be vigilant in a pair. With simple risk allocation, the ESS solution is zero vigilance. Thus each individual invests in the survival of its irreplaceable partner because having a partner increases its own survival. Potentially, this is a widespread phenomenon for animals in the nonbreeding season because animals lost to predation cannot be replaced until the next breeding season. It will only have a large effect, however, when loss of another individual strongly affects the fitness of a focal individual. This could happen if animals spend the nonbreeding season in small, closed groups. Given that nonbreeding groups are often fluid, mutual dependence might apply more widely to mated pairs that stay together year-round. If mates are difficult to replace or mating increases with experience, then each member of the pair would do well to ensure that its partner survives.

Thus individuals might often be "mate guarding" but in the broad sense of protecting their partner's survival and ability to reproduce. Perhaps mutual dependence might contribute to the year-round territorial behavior of tropical birds.

CONCLUSIONS

In this entry the initial models of vigilance were very simple and yet considered multiple attacks. This exercise was very successful because it was able to re-create results from previous models while showing that these results are not general to multiple attacks. Extending these models to partial information worked well, though general and tractable models were only found for pairs of individuals and not for larger groups. For risk allocation and mutual dependence, the rationale for the first models is limited. The models for risk allocation developed here provide a baseline for analyzing mutual dependence. Here again, the models of mutual dependence were neither as simple nor as general as the models of vigilance they build upon, and other models might more readily demonstrate mutual dependence. Thus simplifying assumptions for one purpose may complicate extending the model to another purpose. To explain the logic of potential contributing factors, it can be very helpful to examine a series of very simple models, each resting on its own foundations. When examining the relative importance of and interactions between contributing factors, more complicated models may be necessary. Simple and complex models inform each other, and a range of tools can be useful on any modeling job.

SEE ALSO THE FOLLOWING ARTICLES

Behavioral Ecology / Cooperation, Evolution of / Evolutionarily Stable Strategies / Foraging Behavior / Game Theory / Predator–Prey Models

FURTHER READING

Bednekoff, P. A. 2001. Coordination of safe, selfish sentinels based on mutual benefits. *Annales Zoologici Fennici* 38: 5–14.

Bednekoff, P. A. 2007. Foraging in the face of danger. In D. W. Stephens, J. S. Brown, and R. C. Ydenberg, eds. *Foraging*. Chicago: University of Chicago Press.

Bednekoff, P. A., and S. L. Lima. 1998. Randomness, chaos, and confusion in the study of anti-predator vigilance. *Trends in Ecology & Evolution* 13: 284–287.

Bednekoff, P. A., and S. L. Lima. 1998. Re-examining safety in numbers: Interactions between risk dilution and collective detection depend upon predator targeting behavior. *Proceedings of the Royal Society B: Biological Sciences* 265: 2021–2026.

Bednekoff, P. A., and S. L. Lima. 2004. Risk allocation and competition in foraging groups: reversed effects of competition if group size varies under risk of predation. *Proceedings of the Royal Society of London Series B: Biological Sciences* 271: 1491–1496.

Lima, S. L. 1990. The influence of models on the interpretation of vigilance. In M. Bekoff and D. Jamieson, eds. *Interpretation and explanation in the study of animal behavior: vol. 2. explanation, evolution, and adaptation.* Boulder, CO: Westview Press.

Lima, S. L., and P. A. Bednekoff. 1999. Temporal variation in danger drives antipredator behavior: the predation risk allocation hypothesis. *American Naturalist* 153: 649–659.

Parker, G. A., and P. Hammerstein. 1985. Game theory and animal behaviour. In P. J. Greenwood and P. H. Harvey, eds. *Evolution: Essays in honour of John Maynard Smith*. Cambridge, UK: Cambridge University Press.

Pulliam, H. R. 1973. On the advantages of flocking. *Journal of Theoretical Biology* 38: 419–422.

Treisman, M. 1975. Predation and the evolution of gregariousness. I. Models for concealment and evasion. *Animal Behaviour* 23: 779-800.

ADAPTIVE DYNAMICS

J. A. J. METZ

International Institute for Applied Systems Analysis

Laxenburg, Austria

Adaptive dynamics (AD) is a mathematical framework for dealing with ecoevolutionary problems that is primarily based on the following simplifying assumptions: clonal reproduction, rare mutations, small mutational effects, smoothness of the demographic parameters in the traits, and well-behaved community attractors. However, often the results from AD models turn out to apply also under far less restrictive conditions. The main AD tools are its so-called canonical equation (CE), which captures how the trait value(s) currently present in the population should develop over evolutionary time, and graphical techniques for analyzing evolutionary progress for one-dimensional trait spaces like pairwise invasibility plots (PIPs) and trait evolution plots (TEPs). The equilibria of the CE, customarily referred to as evolutionarily singular strategies (ess's), comprise, in addition to the evolutionary equilibria (or ESSs), points in trait space where the population comes under a selective pressure to diversify. Such points mathematically capture the ecological conditions conducive to adaptive (Darwinian) speciation.

CONTEXT

Micro-, Meso-, and Macroevolution

Adaptive dynamics was initiated as a simplified theoretical approach to mesoevolution, defined here as evolutionary changes in the values of traits of representative individuals and concomitant patterns of taxonomic diversification. This is in contrast to microevolution, i.e., changes in gene frequencies on a population dynamical

time scale, and macroevolution, a term that then can be reserved for changes like anatomical innovations, where one cannot even speak in terms of a fixed set of traits.

Mesoevolution is more than microevolution writ large, and a similar statement applies to macro- versus mesoevolution. Each of these levels has its own emergent phenomena and its own explanatory frameworks, which in the end should be based on idealized summaries of the outcomes of lower-level mechanisms. Trait changes result from the microevolutionary process of mutant substitutions taking place against the backdrop of a genetic architecture and developmental system as deliverers of mutational variation, internal selection keeping the machinery of a body in concert, and ecological selection due to the interactions of individuals with their conspecifics, resources, predators, parasites, and diseases. AD focuses on these encompassing mechanisms.

To get a clean story, AD assumes a time-scale separation between ecology and evolution. In reality, this assumption holds only rather rarely. The idea is that arguments based on it may often lead to outcomes that are fair approximations, provided one applies them selectively and takes a sufficiently gross look at reality. The time-scale argument considerably eases the transition from population genetics to the perspective of ecologists, morphologists, and taxonomists. AD aims at addressing that larger picture at the cost of being wrong in the details.

The Fitness Landscapes of Mesoevolution

Mesoevolution proceeds by the selective filtering by the ecology of a continual stream of mutations. AD concentrates on the ecological side of this process, as at that end there are clearer a priori mathematical structures. The basic theory assumes clonal reproduction, and only a subset of the results extends to the Mendelian case—for monomorphic populations directly and for polymorphic populations after appropriate modification.

When approaching the evolutionary process from ecology, one sees immediately that fitnesses are not given quantities. They depend not only on the traits of individuals but also on their environment. The ecological feedback loop means that in the monomorphic and clonal cases necessarily the fitnesses of types permanently present on the ecological time scale are zero. Only the fitnesses of potential mutants can be positive or negative. The signs and sizes of these mutant fitnesses determine the direction and speed of evolution. Evolution corresponds to uphill movement in fitness landscapes that keep changing so as to keep the resident types at zero (see Fig. 1).

The main insight from the mathematical analysis of this picture has been the discovery of a potential mechanism for adaptive speciation that appears with a certain ubiquity in ecological models. Apart from that, the theory has produced effective tools for analyzing special families of ecoevolutionary models.

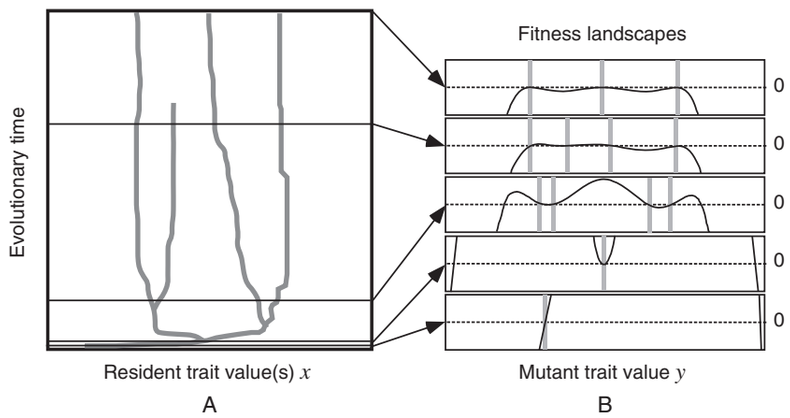


FIGURE 1 (A) Evolutionary path simulated on the basis of a population dynamical model, assuming clonal reproduction. Only the traits that are dominantly present in the population are shown. The second ascending branch finishes since its subpopulation went extinct. (B) The fitness landscapes for five population compositions as these occurred at the indicated times. The vertical bars indicate the types that at that moment were present in the population. At the second selected time, the population resided at a branching point. At the final time, the remaining three subpopulations reside at an evolutionarily steady coalition.

In the section that follows, first clonal reproduction is assumed, where relevant followed with a discussion of the Mendelian case. In line with the landscape analogy, zero fitness is referred to as sea level, etc.

PRELIMINARIES

Fitness

HISTORY

The concept of fitness as a quantitative measure of competitive prowess is a recent invention. Darwin never used the term in this way, and neither did the founding fathers of population genetics. (Except in one of Fisher's early papers; elsewhere Fisher, Haldane, and Wright use terms like selective advantage.) In modern population genetics, fitness is generally seen as the probability to survive to reproduction. However, this only works for relatively simple ecological scenarios with the different life phases neatly separated and synchronized. In ecology, one has to account for a less simple world where populations have age, size, spatial, and other structures and where demographic properties vary with the weather and local conditions.

POPULATION STRUCTURE AND EVOLUTIONARY ENVIRONMENTS

In AD, anything outside an individual that influences its population dynamical behavior (which by definition consists of impinging on the environment, giving birth, and dying) is called environment. It is then always possible in principle to find a representation of that behavior in terms of a state space, transition probabilities that depend on the course of the environment, and outputs that depend on the state of the individual and the condition of the environment at the time. Given the course of the environment, individuals independently move through their state spaces, the population state is a frequency distribution over this space, and the mathematical expectation of this frequency distribution, which is again a frequency distribution, moves according to what mathematicians call a positive linear dynamics (linear due to the independence, positive since we cannot have negative numbers of individuals). Mathematics then says that generally the expected size of a population in an ergodic environment will in the long run on average grow or decline exponentially. (The technical term ergodic means roughly that the environment may fluctuate but that these fluctuations have no persistent trend.) This growth rate, ρ, is what ecologists call fitness. It is necessarily a function of two variables—the type of the individuals parameterized by their traits, Y, and the environment, E, written as $\rho(Y \mid E)$, pronounced as "the fitness of Y when the environment is E" (the vertical bar as separator of the arguments is a notation borrowed from probability theory). The mathematical theory of branching processes (a mathematical discipline that deals with independently reproducing objects), moreover, says that a population starting from a single individual will, barring some technical conditions, either eventually go extinct or grow exponentially, with the probability of the latter being positive if and only if its fitness is positive. For constant environments, ρ is usually written as r and referred to as intrinsic rate of natural increase or Malthusian parameter.

In the theory of longer-term adaptive evolution, one is only interested in populations in which the number of individuals exposed to similar environments is sufficiently large that the internal workings of these populations can be modeled in a deterministic manner. In nature, populations are necessarily bounded. A mathematical consequence is that the community dynamics will converge to an attractor, be it an equilibrium, a limit cycle, or something more complicated. The corresponding environments are not necessarily ergodic, but the exceptions tend to be contrived. So here it will be assumed that community attractors generate ergodic environments.

Let the environment generated by a coalition of clones $C = \{X_1, \ldots, X_n\}$ be written as $E_{\text{attr}}(C)$. Then $\rho(Y \mid E_{\text{attr}}(C))$ is the invasion fitness of a new type Y in a C-community.

It follows immediately that all residents, i.e., types that are present in a community dynamical attractor, have zero fitness, since resident populations by definition do not in the long run grow or decline (see Fig. 1). This fact is basic to the following considerations. For ease of exposition, it is assumed throughout that $E_{\text{attr}}(C)$ is unique; the main results extend to the general case with small modifications.

MENDELIAN DIPLOIDS

The extension of the previous framework to Mendelian populations is easier than perhaps expected (although implementing it can be difficult). For the community dynamics, one has to distinguish individuals according to their genotypes and incorporate their mating opportunities with different genotypes into the description of the environment (in the case of casual matings, with pair formation it becomes necessary to extend the state space of individuals to keep track of their marriage status). Alleles reproduce clonally and as such have invasion fitnesses. It is also possible to define a (mock) fitness of phenotypes

by introducing a parallel clonal model with individuals passing through their lives like their Mendelian counterparts and having a reproduction equal to the average of the contributions through the micro- and macrogametic routes (for humans, semen and ova). With such a definition, some essential, but not all, fitness-based deductions for the clonal case extend to the case of Mendelian inheritance. In particular, for genetically homogeneous populations a resident also has fitness zero (since genetically homogeneous populations breed true.) Moreover, the invasion of a new mutant into a homogeneous population is correctly predicted, as that mutant initially only occurs in heterozygotes that breed true by backcrossing with the homogeneous resident.

Mesoevolutionarily Statics: ESSs

Evolution stays put whenever the community produces an environment such that all mutants that differ from any of the residents have negative fitness. In the special case of a single resident type, we speak of an evolutionarily steady strategy (ESS). (The old name evolutionarily stable strategy is a bit of a misnomer since ESSs need not be evolutionarily attractive.) The general case when there may be more than one resident type is called an evolutionarily steady coalition (ESC). ESCs are the equilibria of evolution.

One way of calculating ESSs is depicted in Figure 2. For each environment as generated by a possible resident, the maximum of the invasion fitness landscape $\rho(Y \mid E_{\text{attr}}(X))$ is calculated. Next, one intersects the resulting curve (or surface) $Y = Y_{\text{opt}}(X)$ with the line (or surface) $Y = X$ to get the ESS $X^* = Y^*$. As monomorphic residents have fitness zero, all potential mutants $Y \neq Y^*$ have negative fitness.

The situation for ESCs is more complicated, as there may be so-called genetic constraints, with heterozygote superiority as generic example. However, in the so-called ideal free (IF) case all phenotypes comprising an ESC have fitness zero (at least when there is a single birth state and the ESC engenders a community dynamical equilibrium). This IF case is defined by the requirement that there are no genetic constraints whatsoever, that is, mutants can occur that produce any feasible type as heterozygotes in the genetic background of the resident population.

Fitness Proxies

The existence of a well-defined fitness forms the basis for the calculation of ESSs and AD. However, given its existence it is often possible to replace ρ by some more easily determined quantity that leads to the same outcome for the calculations of interest. For example, in

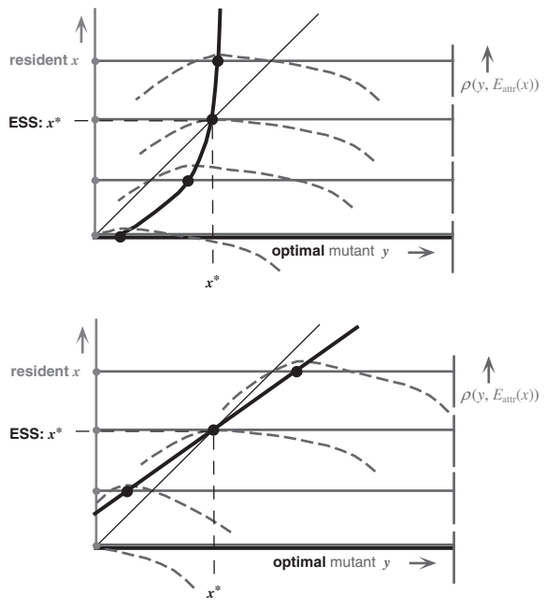


FIGURE 2 Scheme for calculating ESSes: For each of the possible resident populations, here characterized by a scalar trait, the invasion fitness of all potential mutants is calculated (interupted curves). The mutant axis is drawn on the same scale as the resident axis. From these fitness curves, the optimal strategy for the corresponding resident environment is calculated (fat curve). The ESS is the optimal reply to itself, to be calculated by intersecting the fat curve with the 45° line. In (A), the ESS attracts evolutionarily as can be seen from the fact that for other trait values the fitness landscape increases in the direction of the ESS. In (B), the fitness landscapes decrease in the direction of the ESS. Hence it repels.

optimization calculations ρ can be replaced with any quantity monotonically related to it, and for the graphical methods of AD one may replace ρ with any sign-equivalent quantity. Such quantities are referred to as fitness proxies. An example of a fitness proxy of the first type is the average rate of energy intake. Being a fitness proxy is always predicated on additional assumptions. For instance, it may help a forager little to increase its energy intake in an environment where this drastically increases its exposure to predation. A fitness proxy of the second type, restricted to nonfluctuating environments, is the logarithm of the average lifetime offspring number $\ln(R_0)$. If individuals may be born in different birth states, as is the case for spatial models, where birth position is a component of the birth state, R_0 is defined as the dominant eigenvalue of the next generation matrix (or operator in the case of a continuum of birth states). This matrix is constructed by calculating from a model for the behavior of individuals how many offspring are born on average in different birth states dependent on the birth state of the parent.

Optimization Principles

Often biologists try to predict evolutionary outcomes by pure optimization. Of course, if one measures only the environment one may predict the evolutionarily steady trait values that go with that environment by maximizing fitness in that environment. This is why predictions from optimization work. In general, optimization procedures do not predict the outcome of evolution, for that entails also predicting the environment that goes with the ESS, but they often satisfactorily predict the strategies that go with that environment. However, such limited predictions are of little use when considering the consequences of environmental change like increasing fishing mortality or global warming.

Many papers on ESS theory also derive optimization principles by which ESSs may be calculated. Hence, it is of interest to know when there exist properties of phenotypes that are maximized at an ESS. It turns out that this is the case if and only if one of two equivalent conditions holds good: the effect of the trait (environment) can be summarized in a single variable such that for each environment (trait) there exists a single threshold above which fitness is positive. These statements can be paraphrased as follows: the trait (environment) should act in an effectively one-dimensional monotone manner. (Think, for example, of the efficiency of exploiting a sole limiting resource.) Given such a one-dimensional summary $\psi(E)$ of the environment (which is often more easily found), it is possible to construct a matching summary of the traits $\phi(X)$, and vice versa, through

$$\phi(X) = -\psi(E_{\text{attr}}(X)) \tag{1}$$

The previous statements hinge on the interpretation of the term optimization principle. The latter should be interpreted as a function that attaches to each trait value a real number such that for any constraint on the traits the ESSs can be calculated by maximizing this function. This proviso mirrors the practice of combining an optimization principle derived from the ecology with a discussion of the dependence of the evolutionary outcome on the possible constraints.

If an optimization principle ϕ exists, each successful mutant increases $\phi(X)$ and hence any ESS attracts. Moreover, $\psi(E_{\text{attr}}(X))$ decreases with each increase in $\phi(X)$. Since fitness increases with ψ where it counts, i.e., around zero, ψ may be dubbed a pessimization principle. When a pessimization principle exists, in the end the worst attainable world remains, together with the type(s) that can just cope with it.

Optimization principles come closest to the textbook meaning of fitness, which generally fails to account for the fact that the fitnesses of all possible types are bound to change with any change in the character of the residents. However, optimization principles, although frequently encountered in the literature, are exceptions rather than the rule.

In evolutionary ecology textbooks, the maximization of r or R_0 takes pride of place without mention of the reference environment in which these quantities should be determined. Hence, all one can hope for is that the same outcome results for a sufficiently large collection of reference environments. It turns out that this is the case if and only if r can be written as $r(X \mid E) = f(r(X \mid E_0), E)$ respectively. $\ln(R_0)$ can be written as $\ln(R_0(X \mid E)) = f(\ln(R_0(X \mid E_0)), E)$ for some function f that increases in its first argument, where E_0 is some fixed but otherwise arbitrary environment. These two criteria are relatively easy to check. A fair fraction of textbook statements, if taken literally, apply only when these special conditions are met. This happens, for instance, when the only influence of the environment is through an additional state- and type-independent death probability, or rate, respectively, when the life history can be decomposed into stages that are entered through single states (so that no information carries over), of which one comprises all states after the onset of reproduction, and no stage is affected by both X and E.

ADAPTIVE DYNAMICS (AD)

Traits, PIPs, MIPs, and TEPs

Paleontologists and taxonomists are interested in the change of traits on an evolutionary time scale. What are traits to taxonomists are parameters to ecologists. So in AD one is after a dynamics in the parameter space of a community dynamics. The trick to arriving at such a simple picture is to assume that favorable mutants come along singly after a community has relaxed to an attractor. Another trick is to assume clonal reproduction, on the assumption that this way one can find out where the ecology would drive evolution if the latter were not hampered by the constraints of Mendelian genetics.

To get at a purely trait-oriented picture, any reference to the environment should be removed from the expression for invasion fitness:

$$s(Y \mid C) = \rho(Y \mid E_{\text{attr}}(C)) \qquad (2)$$

(often this is written as $s_C(Y)$ to emphasize the interpretation as a family of fitness landscapes).

This subsection focuses on scalar traits, starting with the case where there is only a single clonally reproducing resident, $C = x$. The first step in the analysis is plotting a contour plot of $s(y|x)$. Usually this is simplified to plotting only the zero contours, as those matter by far the most. The result is called a pairwise invasibility plot (PIP) (see Fig. 3).

Note that the diagonal is always a zero contour, as residents have fitness zero. A point where some other contour crosses the diagonal is referred to as an evolutionarily singular strategy (ess). The ESSes are a subset of the ess-es.

Now assume that mutational steps are small and that in the beginning there is only one resident trait value $x(0)$. Plot this value on the abscissa of the PIP, say, in the left panel of Figure 3. After some random waiting time, mutation creates a new trait value y. This trait value can invade only when it has positive fitness, i.e., is in a plus area of the PIP. It can be proved that an invading type replaces its progenitor if the latter is not too close to an ess or a bifurcation point of the community dynamics and the mutational step is not too large. If such a replacement has occurred, we call the new trait value $x(1)$. In the PIP under consideration, if $x(0)$ lies to the left of the ESS, then $x(1)$ lies to the right of $x(0)$, and vice versa. Hence, repeating the process pushes the evolutionary path to a neighborhood of the ESS. Upon reaching that neighborhood, it may become possible that ecologically the mutant and its progenitor persist together.

To see how such coexisting pairs of strategies fare evolutionarily, it is necessary to consider the set of protected dimorphisms, i.e., pairs of strategies that can mutually invade, denoted as (x_1, x_2). Its construction is shown in Figure 4.

The evolutionary movement of the pair (x_1, x_2) is governed by $s(y \mid x_1, x_2)$. Under the assumption of small mutational steps, a good deal of information can be extracted from the adaptive isoclines, calculated by setting the selection gradient

$$g_i(x_1, x_2) = \mathrm{d}s(y \mid x_1, x_2)/\mathrm{d}y\big|_{y=x_i} \qquad (3)$$

equal to zero. As depicted in Figure 5, x_1 will move to the right when g_1 is positive and to the left when it is negative, and x_2 will move up when g_2 is positive and down when it is negative.

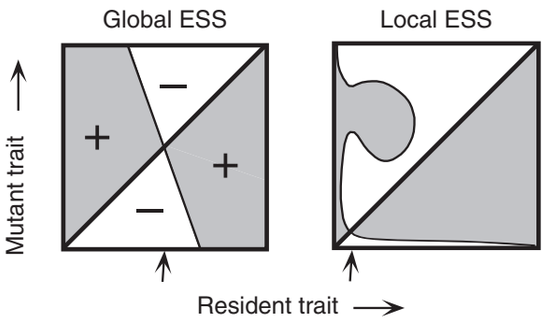


FIGURE 3 Pairwise invasibility plots: sign of the fitness of potential mutants as a function of the mutant and the resident traits.

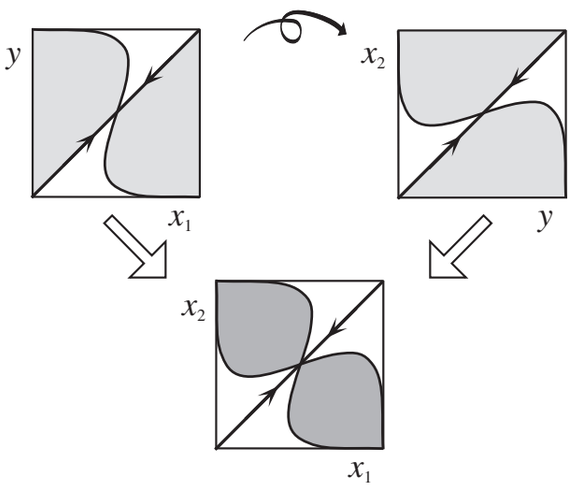


FIGURE 4 The construction of a *mutual invasibility plot* (MIP), depicting the set in (trait space)^2 harboring protected dimorphisms. Not all polymorphisms occurring in AD are protected, but unprotected polymorphisms have the habit of never lying close to a diagonal.

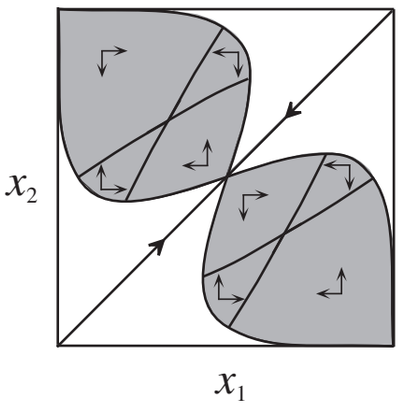


FIGURE 5 *Trait evolution plot* (TEP), i.e., MIP together with arrows that indicate the direction of the small evolutionary steps that result from the invasion by mutants that differ but little from their progenitor, and adaptive isoclines.

From the classification of the possible dynamics near an ess in Figure 7, it can be seen that the ess in the PIPs from Figure 3 also attracts in the dimorphic regime.

The Canonical Equation

BASICS

The dependence of the dynamical outcomes on no more than the sign of the invasion fitness hinges on the ordering properties of the real line. For vectorial traits, one must proceed differently. The workhorse is the so-called canonical equation (CE) of AD, a differential equation that captures how the trait vector changes over evolutionary time, on the assumption of small mutational steps, accounting for the fact that favorable mutants do not always invade due to demographic fluctuations. For unbiased mutational steps, the evolutionary speed and direction are given as the product of three terms: from left to right, (1) the effective population size N_e (as in the diffusion equations of population genetics), (2) the probability of a mutation per birth event ε times a matrix C consisting of the variances and covariances of the resulting mutational step, and (3) the selection gradient (the "curly d" notation stands for differentiating for that variable while treating the other variables as parameters):

$$G(X) = \left. \begin{pmatrix} \partial s(Y \mid X)/\partial y_1 \\ \vdots \\ \partial s(Y \mid X)/\partial y_n \end{pmatrix} \right|_{Y=X}, \qquad (4)$$

which is a vector pointing from the position of the resident in the steepest uphill direction. The uphill pull of selection is thus modified by the differential directional availability of mutants expressed in the mutational covariance matrix. For diploid Mendelian populations, an additional factor 2 appears, since the substitution of a mutant allele leads to a mutant homozygote twice as far removed (at the considered order of approximation). The equilibrium points of the CE are the ess-es mentioned previously.

In reality, many mutants attempt to substitute simultaneously. Luckily, for small mutational steps this appears to affect the environment only in the higher-order terms that in the derivation of the CE disappear. However, in the clonal case the effects of the mutants do not add up, since a mutant may be supplanted while invading by a better mutant from the same parent type. In the Mendelian case, the CE will do a better job as substitutions occur in parallel on different loci, which to the required order of approximation act additively.

THE LINK WITH EVO-DEVO

From an AD perspective, the link with evolutionary developmental biology (Evo-Devo) is through the mutational covariance matrices. Unfortunately, Evo-Devo has yet little to offer in this area. Therefore, at present the most that AD researchers can do is work out how the outcomes of a specific ecoevolutionary model depend on the possible forms of the mutational covariance matrix. The answers from AD thus become Evo-Devo questions: is the mutational covariance matrix for these traits expected to fall within this or that class?

To show the importance of the missing Evo-Devo input in AD: mutational covariance matrices have an (often dominating) influence on the time scales of evolution (Fig. 6A) and the basins of attraction of ess-es (Fig. 6B), even to the extent that they often determine whether an ess attracts or not.

On a more philosophical level, it bears noting that the selection gradient points only in a single direction, while the components of the trait vector orthogonal to that gradient hitchhike with the selectively determined motion thanks to a developmental coupling as expressed in the mutational covariance matrix. The higher the dimension of the trait space, the larger the contribution of development as a determinant of evolutionary motion. The dimensions of the trait spaces that are routinely considered thus make for the contrast in attitudes of behavioral ecologists and morphologists, with the former stressing selection and the latter developmental options for change.

Evolutionarily Singular Strategies

Evolutionarily singular strategies X^* can be calculated by setting the fitness gradient equal to zero. Figure 7 shows their classification according to dynamical type for scalar

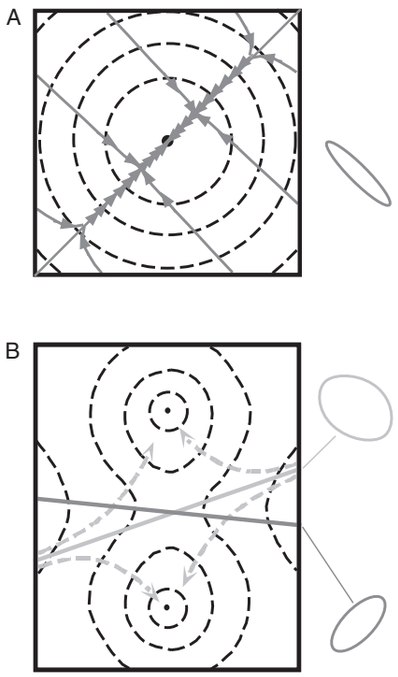


FIGURE 6 Two fitness landscapes that are supposed to keep their shape and sink only when the adaptive trajectory moves uphill (as is the case if and only if the population regulation is through an additional state-independent death rate). Distributions of mutational steps are symbolized by ovals. (A) The shape of the mutation distribution induces a time scale separation between the movement along the diagonal and anti-diagonal direction. (B) The difference in mutation distributions causes a difference in the domains of attraction of the two ESSes.

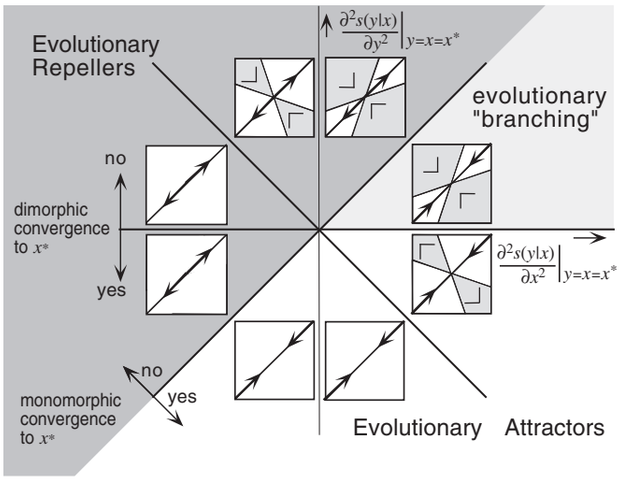


FIGURE 7 A classification of the ESSes for scalar traits. The cases in the lower half are all ESSes. The leftmost of these repels, the others attract. The latter ESSes are thus genuine evolutionary attractors. The branching points in the rightmost upper sector attract monomorphically but repel dimorphically.

traits. (Note that this classification was derived deductively from no more than some mathematical consistency properties shared by well-posed ecoevolutionary models.) Devising a good classification for higher-dimensional ess-es is an open problem. One reason is that in higher dimensions the attractivity of a singular point crucially depends on the mutational covariance matrix except in very special cases.

Adaptive Speciation

Perhaps the most interesting ess-es are branching points, where the ecoevolutionary process starts generating diversity. When approaching such points, the evolutionary trajectory, although continually moving uphill, gets itself into a fitness minimum. More precisely, it is overtaken by a fitness minimum. This is due to the population dynamics; think of the following analogy: Somewhere gold has been found, attracting people to that spot. However, after the arrival of too many diggers it becomes more profitable to try one's luck at some distance.

The buildup of diversity can take very different forms. In the clonal case, the population just splits into two as depicted in Figure 1. In the Mendelian case, the diversification starts with a broadening of the variation in the population. The fitness landscape locally has the shape of a parabola that increases away from x^*. This means that types more on the side have a higher fitness than those in the center. It therefore pays not to beget kids near the center. The Mendelian mixer has the contrary tendency to produce intermediate children from dissimilar parents. Fortunately, there are all sorts of mechanisms that may thwart this counterproductive mixing. The most interesting of these is the buildup of some mechanism that allows like extremes to mate only among themselves, thus ensuring that the branches become separate genetic units. A simple mechanism occurs in insects that diversify in their choice of host plants, with mating taking place on those hosts.

The author's conviction is that in cases where there is no automatic mating barrier, a buildup of other mechanisms engendering assortative mating is not unexpected. Present-day organisms are the product of 3.5 billion years of evolution. During that time, their sensory and signaling apparatus has been evolutionarily honed for finding the most advantageous mates. Hence, there will be an abundance of template mechanisms. These mechanisms, once recruited to the task, will probably have a tendency to enhance each other in their effect. One may thus expect that the available generalized machinery can often

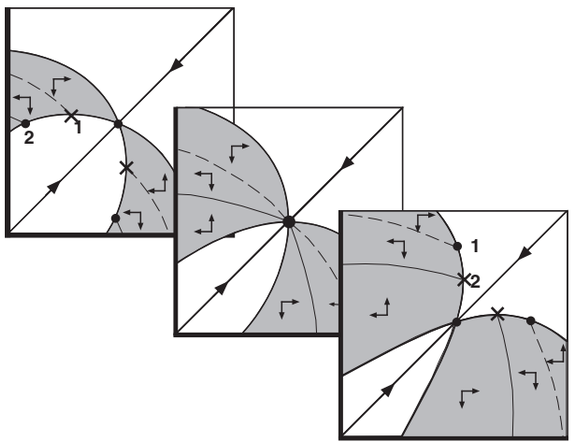


FIGURE 8 Three TEPs corresponding to a bifurcation of an ESS to a branching point. In this case, the adaptive tracking of the ESS in the wake of slow global environmental change stops due to a change in character of the ESS. In the fossil record, this scenario would correspond to a punctuation event that starts with speciation.

easily be adapted so as to genetically separate the branches whenever evolution brings the population to a branching point. However, most scientists working on the genetics of speciation do not appear to agree with this view.

Bifurcations

The bifurcations of AD encompass all the classical bifurcations found in ecological models. In addition, there is a plethora of additional bifurcations. An example is the transition from an ESS to a branching point depicted in Figure 8.

JUSTIFYING THE AD APPROXIMATION

Traits and Genotype-to-Phenotype Maps

The real state space of the mesoevolutionary process is genotype space, while the phenotypic trait spaces of AD are but convenient abstractions. Phenotypic mutational covariances reflect both the topology of genotype space, as generated by mutational distances, and the genotype-to-phenotype map generated by the developmental mechanics. This reflection can only be expected to be adequate locally in genotype space, and therefore locally in evolutionary time.

AD focuses on small mutational steps. A partial mechanistic justification comes from the expectation that the evolutionary changes under consideration are mostly regulatory. Coding regions of genes are in general preceded by a large number of short regions where regulatory material can dock. Changes in these regulatory regions lead to changes in the production rate of the gene product. The influence of a single regulatory region among many tends to be rather minor. Note, moreover, that phenotypes should in principle be seen as reaction norms, i.e., maps from microenvironmental conditions to the characteristics of individuals (another term is conditional strategies). The phenotypes of AD are inherited parameters of these reaction norms. Only in the simplest cases is a reaction norm degenerate, taking only a single value.

Internal Selection

Functional morphologists usually talk in terms of whether certain mechanisms work properly or not and discuss evolution as a sequence of mechanisms all of which should work properly, with only slight changes at every transformational step. Translated into the language of fitness landscapes, this means that only properly working mechanisms give fitnesses in the ecologically relevant range, while the improperly working ones always give very low fitnesses. This leads to a picture of narrow, slightly sloping ridges in a very high-dimensional fitness landscape. The slopes on top of the ridges are the domain of ecology; their overall location is largely ecology independent.

As a simple example, one may think of leg length. The left and right leg are kept equal by a strong selection pressure, which keeps in place a developmental system that produces legs of equal length, notwithstanding the fact that during development there is no direct coupling between the two leg primordia. Hence in a trait space spanned by the two leg lengths, ecologists concentrate on only the diagonal.

The trait spaces dealt with in morphology are very high dimensional so that the top of a ridge may be higher dimensional, while away from the ridge the fitness decreases steeply in a far larger number of orthogonal directions. A picture similar to that of functional morphologists emerges from Evo-Devo. The long-term conservation of developmental units (think of the phylotypic stage or homology) can only be due to strong stabilizing selection, since mutations causing large pattern changes have many side effects with dire consequences. As a result, ecological selection generally acts only on quantitative changes in the shapes and sizes of homologous body parts.

Since the fitness differences considered in functional morphology and Evo-Devo largely result from the requirement of an internal coherence of the body and of the developmental process, people speak of internal selection. Here, the term is used to label features of the fitness landscape that are roughly the same for all the environments that figure in an argument.

The Assumptions of AD

THEORY

Figure 9 illustrates an argument by Fisher showing that the higher the dimensionality of the trait space, the more difficult becomes the final convergence to an adaptive top. This argument extends to the movement in a ridgey fitness landscape: the higher the number of orthogonal off-ridge directions, the more rare it is for a mutational step to end up above sea level, and small mutational steps have a far higher propensity to do so than large ones. Together, these two arguments seem to underpin the requirements of AD that mutations in ecologically relevant directions are scarce and the corresponding mutational steps are small.

Unfortunately the above arguments contain a biological flaw: the assumed rotational symmetry of the distribution of mutational steps. Real mutation distributions may be expected to show strong correlations between traits. Correlation structures can be represented in terms of principal components. The general experience with biological data is that almost always patterns are found like the ones shown in Figure 10B. Figure 10C shows that the existence of mutational correlations will in general enhance the rareness and smallness of the mutational steps that end up above sea level.

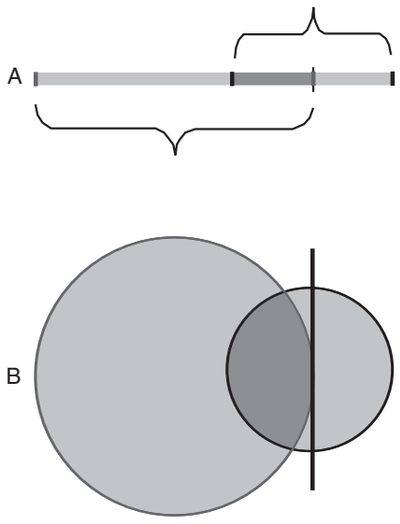


FIGURE 9 (A) Two balls in $\mathbb{R}^1$, with the center of the smaller ball on the boundary of the larger ball. The ratio of the volume of their intersection to the volume of the smaller ball is $\frac{1}{2}$. (B) A similar configuration in $\mathbb{R}^2$. The volume of the intersection is now a smaller fraction of the volume of the smaller ball. For similar configurations in $\mathbb{R}^n$, this fraction quickly decreases to zero for larger n. Now think of the larger ball as the part above sea level of a fitness hill and of the smaller ball as a mutation distribution. Clearly the fraction of favorable mutants will go to zero with n.

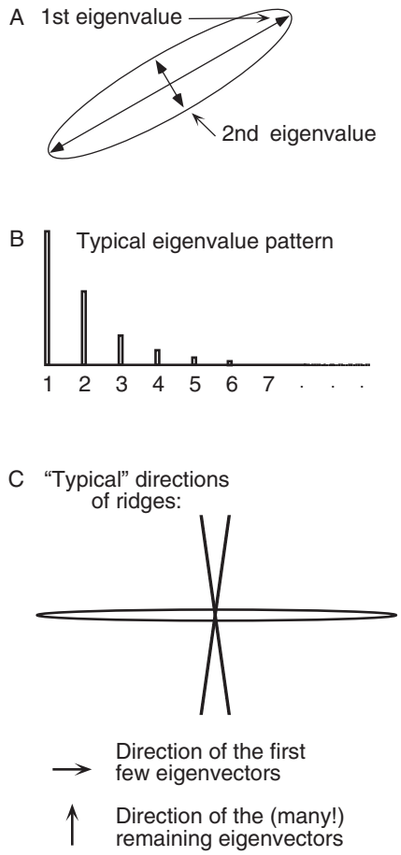


FIGURE 10 (A) Contour line of a bivariate distribution, supposedly of mutational steps. The lengths of the two axes of the ellipse, called principal components, are proportional to the square root of the eigenvalues of the mutational covariance matrix. (B) Typical eigenvalue pattern found for large empirical covariance matrices. (C) The mutation distribution will rarely be fully aligned with the fitness ridges. If in a high dimensional fitness landscape one takes one's perspective from the mutation distribution and looks at the orientation of the ridges relative to the first few principal axes of this the mutation distribution, then, when the number of the dimensions of the trait space is very large and the ridge has a relatively low dimensional top, the ridge will typically extend in a direction of relatively small mutational variation.

DATA

The conclusions from the previous subsection seem to underpin nicely the assumptions of AD. Unfortunately, various empirical observations appear to contradict these conclusions. Populations brought into the lab always seem to harbor sufficient standing genetic variation to allow quick responses to selection, and a few loci with larger effect (quantitative trait loci, or QTL) are often found to underlie the variation in a trait. However, these empirical observations may have less bearing on the issue than one might think.

First, given the speed of evolution relative to the changes in the overall conditions of life, populations in the wild are probably most often close to some ESS. Moreover, in

noisy environments, fitness maxima tend to be flat. This means that near-neutral genetic variation will accumulate, which is exploited first when a population gets artificially selected upon. At ESSes, the mutation limitation question is largely moot. Beyond its statics, AD's main interest is in the larger-scale features of evolutionary trajectories after the colonization of new territory or, even grander, a mass extinction. The scale of these features may be expected to require a further mutational supply of variation. Second, directional selection on an ecological trait may be hampered by stabilizing internal selection on pleiotropically coupled traits. In the lab, this stabilizing selection is relaxed. This means that far more variation becomes available for directional selection than is available in the wild.

Finally, AD-style theory has shown that in the absence of assortative mating the initial increase of variability after the reaching of a branching point tends to get redistributed over a smaller number of loci with increasing relative effect The end effect will be QTL, but it will be produced through the cumulative effect of small genetic modifications.

SEE ALSO THE FOLLOWING ARTICLES

Adaptive Landscapes / Branching Processes / Coevolution / Evolutionarily Stable Strategies / Mutation, Selection, and Genetic Drift

FURTHER READING

Dercole, F., and S. Rinaldi. 2008. *Analysis of evolutionary processes: the adaptive dynamics approach and its applications.* Princeton: Princeton University Press.

Dieckmann, U., M. Doebeli, J. A. J. Metz, and D. Tautz, eds. 2004 *Adaptive Speciation*. Cambridge Studies in Adaptive Dynamics vol. 3. Cambridge, UK: Cambridge University Press.

Dieckmann, U., and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of Mathematical Biology* 34: 579–612.

Diekmann, O. 2004. A beginner's guide to adaptive dynamics. *Mathematical Modelling of Population Dynamics* 63(4): 47–86.

Durinx, M., J. A. J. Metz, and G. Meszéna. 2008. Adaptive dynamics for physiologically structured models. *Journal of Mathematical Biology* 56: 673–742.

Geritz, S. A. H., É. Kisdi, G. Meszéna, and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology* 12: 35–57.

Gyllenberg, M., J. A. J. Metz, and R. Service. 2011. When do optimisation arguments make evolutionary sense? In F. A. C. C. Chalub and J. F. Rodrigues, eds. *The mathematics of Darwin's legacy*. Basel, Switzerland: Birkhauser.

Leimar, O. 2009. Multidimensional convergence stability. *Evolutionary Ecology Research* 11: 191–208.

Metz, J. A. J. 2008. Fitness. In S. E. Jørgensen and B. D. Fath, eds. *Evolutionary ecology.* Vol. 2 of Encyclopedia of Ecology. Oxford: Elsevier.

Metz, J. A. J. 2011. Thoughts on the geometry of meso-evolution: collecting mathematical elements for a post-modern synthesis. In F. A. C. C. Chalub and J. F. Rodrigues, eds. *The mathematics of Darwin's legacy.* Basel, Switzerland: Birkhauser.

Metz, J. A. J., S. A. H. Geritz, G. Meszéna, F. J. A. Jacobs, and J. S. van Heerwaarden. 1996. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In S. J. van Strien and S. M. Verduyn Lunel, eds. *Stochastic and spatial structures of dynamical systems*. Amsterdam: North-Holland.

ADAPTIVE LANDSCAPES

MAX SHPAK

University of Texas, El Paso

The concept of an adaptive (or fitness) landscape was introduced by Sewall Wright as a metaphor for the genetic "space" on which adaptive evolution occurs. The metaphor also drew on analogies from the physical sciences, in that the adaptive landscape was conceptualized as a sort of inverted potential energy surface. Essentially, the view that evolution takes place on an adaptive landscape entails representing natural selection as deterministic "hill climbing"—with the peaks representing local fitness optima and valleys representing fitness minima. The metaphor proved to be an enduring one, principally because of its intuitive visual appeal. However, it has also caused a great deal of confusion, in part because of its use to describe two entirely different entities, one of which is dynamically sufficient for a description of natural selection, the other only under certain limiting assumptions (the first being the space of possible genotypes; the second, mean population fitness as a function of allele frequencies). Furthermore, it will be argued that excessive reliance on the three-dimensional metaphor of a rugged landscape results in flawed analyses that do not hold up to more formal treatments of biologically realistic multilocus model systems.

THE SELECTION EQUATIONS

In order to make sense of adaptive landscapes, we first consider the selection equations. For the sake of notational simplicity and without loss of generality, we begin with a population of haploid asexual organisms, starting with the standard scenario of a single locus with two alleles in an effectively infinite population. Denoting the frequency of one allele as p and the other as $1 - p$, we assign respective fitness values (defined as viability, fecundity, or the product of the two) to be W_1, W_2. The mean population fitness is $\overline{W} = pW_1 + (1 - p)W_2$, while the expected change in the frequency of the first allele in discrete time is

$$E[\Delta p] = \frac{p(1-p)}{\overline{W}}(W_1 - W_2) \qquad (1)$$

where $\Delta p = p(t + 1) - p(t)$.

For instance, when $W_1 > W_2$, Equation 1 states that the frequency of the first allele increases in proportion to the fitness difference between it and the competing allele. This equation is a special case of a more general expression for the change in the frequency of any true-breeding trait value x, i.e.,

$$E[\Delta x] = \frac{var(x)}{\overline{W}} \frac{\partial \overline{W}}{\partial x} \quad (2)$$

where var(x) is the variance of the trait value and the derivative measures the change in mean fitness with respect to the heritable trait. Equation 2 provides the basis for one concept of the adaptive landscape, i.e., the mean population fitness $\overline{W}(x)$ as a function of allele or trait frequency in the population. In part, this interpretation was motivated by Fisher's fundamental theorem, which states that in the absence of frequency dependence or the confounding effects of genetic linkage, the mean population fitness increases at a rate proportional to its variance; i.e.,

$$\Delta \overline{W} = \frac{var(\overline{W})}{\overline{W}} \quad (3)$$

Furthermore, in a finite population subject to genetic drift as well as natural selection, we can approximate the probability density of the allele frequency at equilibrium as

$$F(p) \approx \overline{W}(p)^N \quad (4)$$

These results underscore the importance of the mean fitness in evolution and lead Wright to conceptualize the fitness landscape in terms of mean population fitness rather than as the fitness values of individual alleles or genotypes. This distinction is illustrated for the simple diallelic diploid example in Figures 1A and 1B, where the first figure plots representative fitness values for the two alleles and the second shows mean population fitness as a function of the frequency of the first allele.

The principal difficulty with the mean fitness representation is that in more realistic evolutionary models that incorporate frequency dependence or mutation, $\overline{W}$ is not necessarily maximized by natural selection. Furthermore, if we include the complications of linkage and recombination, we no longer generally have a dynamically sufficient description of $\overline{W}(p_1 \ldots p_n)$ as a function of allele frequencies at each locus $1 \ldots n$, because the marginal fitness of an allele depends on which particular allelic states it is associated with at other loci.

In contrast, one can always derive a dynamically sufficient description of selection (analogous to Eq. 1) from the fitness values of individual genotypes, either for the individual fitness values or for mean population fitness as a function of genotype frequency. Much of the confusion in the interpretation of adaptive landscapes arose from confounding the two representations. Unless otherwise indicated, this entry will use the term adaptive landscape to refer to the genotype space representation. It will be argued that this form is often more fruitful in addressing a number of questions relevant to our understanding of adaptation, speciation, and in bridging the conceptual divide between macro and microevolution.

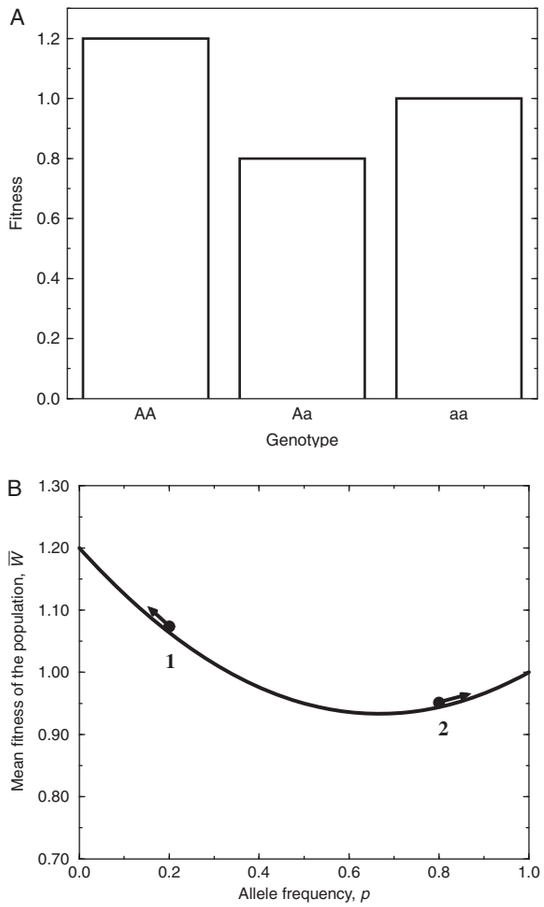


FIGURE 1 (A) The adaptive landscape as a function of genotypes, in this case for a single diploid locus. (B) The adaptive landscape as a function of genotype frequency, based on the fitness values of individual diploid genotypes in part (A).

MUTATION AND MULTIPLE LOCI

Understanding fitness functions and the structure of adaptive landscapes for multilocus genotypes requires an increase in the dimensionality of the model system. This is particularly the case if sexual reproduction, and consequently recombination, introduce additional dynamical complexity into the models.

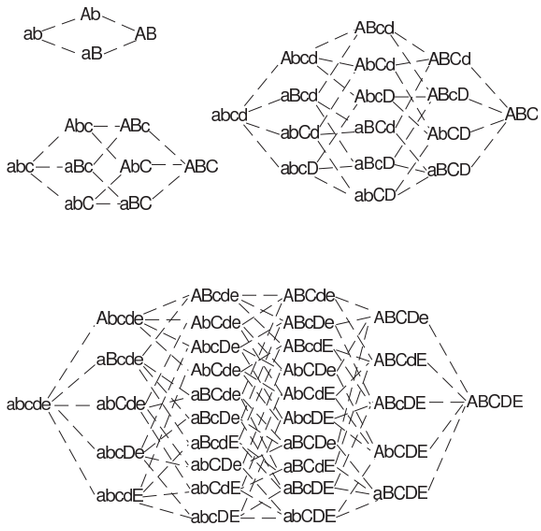


FIGURE 2 Graphs in haploid genotype space for n = 2, 3, 4, and 5 loci and 2 alleles, illustrating the increase in the number of paths between any two genotypes as a power function of the number of loci.

Begin by considering the simplest scenario—a haploid asexual organism with n loci and 2 alleles (results are generalizable to a 4-letter nucleotide or 20-letter amino acid alphabet, so a binary model is chosen for the sake of simplicity). In this model, there are 2^n possible genotypes, and if we assign a fitness value W_i to the ith genotype, we can describe the changes in their frequency with a trivial modification of Equation 1. However, a complete model of evolution requires both a selection and a transmission operator, where the latter involves mutation and/or recombination. The transmission operator imposes a topology onto the space of 2^n genotypes by defining a neighborhood of each genotype x_i. For example, if mutations are rare, a first-order approximation assumes that only single-point mutations are possible. Representing every genotype as a point on a lattice with n neighbors, the genotype space defined by point mutation is a hypercube, shown in two through five dimensions, corresponding to 2–5 loci, in Figure 2.

If we allow multiple mutations, the graph's connectivity increases, up to the case where every vertex has an edge with all $2^n - 1$ others. In a model life cycle where selection occurs before reproduction and mutation, the dynamics on a general fitness landscape are defined by the change in the frequency x_i of every genotype:

$$x_i(t+1) = \frac{1}{\overline{W}}\left(w_i x_i(t) + \sum_{j \neq i}(\mu_{ji} w_j x_j(t) - \mu_{ij} w_i x_i(t))\right) \tag{5a}$$

where μ_{ij} is the probability of mutating from state i to state j in a single time interval. If one interprets x as absolute rather than relative frequencies, we can write Equation 5a without the mean fitness term. In this case, Equation 5 becomes a system of linear equations involving the mutation and selection operators, i.e.,

$$\vec{x}(t+1) = A\vec{x}(t) \tag{5b}$$

for $\mathbf{x}$ a vector of genotype absolute frequencies, and $\mathbf{A} = \mathbf{MW}$, where $\mathbf{M}$ is a matrix of mutation probabilities (i.e., for point mutations, $M_{ii} = (1 - n\mu)$, $M_{ij} = \mu$ if the Hamming distance between i and j is equal to unity, otherwise $M_{ij} = 0$) and $\mathbf{W}$ is a diagonal matrix with genotype fitness $W_{ii} = W_i$.

In an infinite population, one can calculate the mutation–selection equilibrium distribution of genotype frequencies from Equation 5b as the dominant eigenvector of $\mathbf{A}$. This distribution is often referred to as the quasispecies, following the terminology of Manfred Eigen and Peter Schuster, who derived the results for model systems of self-replicating macromolecules.

For a complete dynamical characterization, the system of equations requires as many variables as there are genotypes. Because the number of genotypes increases as a power of the number of loci, finding analytical solutions becomes intractable because of the high dimensionality. However, if one assumes certain symmetry properties in the fitness function W_i, the description of mutation–selection dynamics can be greatly simplified. We begin by considering some canonical examples of "simple" fitness landscapes—the additive fitness function and the flat (neutral) landscape, before proceeding to more general models.

Additive Landscapes

The additive landscape is so called because it is assumed that the fitness contribution of an allele at every locus is independent of the allelic state at other loci, so that the fitness of any genotype is just the sum (or product) of the value of each allele, i.e., $W_i = s_1 + s_2 \cdots + s_n$, where s_k is the fitness contribution of allelic state s at locus k. This generates a landscape with a single optimum or fitness peak, as shown in Figure 3.

Biologically, one can interpret the single-peak landscape as a scenario where there is one most fit genotype and that every point mutation away from it decreases the fitness incrementally. If we make the further assumption that selection is relatively weak, higher-order interaction terms across loci can be ignored, so that the contributions from each site will be approximately additive, which

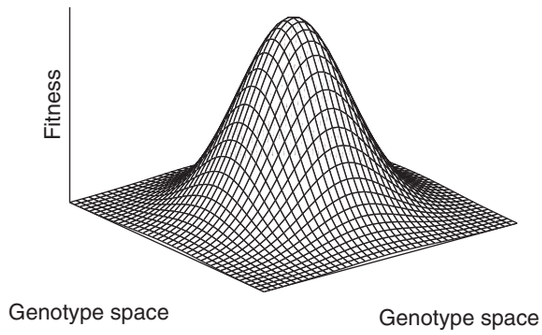


FIGURE 3 An additive fitness landscape in genotype space with a single optimum, under the assumption that each point mutation away from the local optimum proportionately decreases genotype fitness.

simplifies the representation of selection dynamics still further.

Quantitative Traits

A special case of an additive fitness landscape involves stabilizing selection on an additive quantitative trait. If we assume that the number of loci n is large and the phenotypic/fitness contributions of alleles are additive and identically distributed random variables at every locus, then we can approximate the phenotype x by a Gaussian distribution with a mean at the optimum (rescaled to $x_{opt} = 0$) and a genetic variance σ^2 determined by the sum of the variance contributions from individual loci. Standard models of stabilizing selection postulate a quadratic or a Gaussian selection function about the mean value, i.e.,

$$W(x) = Exp\left[-\frac{x^2}{2V}\right]$$

where V is a variance parameter that measures the intensity of stabilizing selection (i.e., small V indicating strong selection against nonoptimal phenotypes). Combining selection and mutation under the assumption of weak mutational phenotypic effects at each locus with an approximately normal distribution, the selection equations on the mean phenotype $\bar{x}$, rescaled with an optimum at 0, are

$$E[\Delta\bar{x}] = -\frac{{\sigma_A}^2}{{\sigma_A}^2 + {\sigma_B}^2 + V}\bar{x} \qquad (6)$$

where ${\sigma_A}^2$ is the additive variance due to mutation and ${\sigma_B}^2$ is the variance due to environmental noise. For larger mutational variances, the normal approximations break down and results based on mixtures of distribution, such as the "house of cards" model, need to be introduced.

Neutral Landscapes

A flat or neutral landscape is simpler still, in that every genotype, regardless of dimensionality, is assumed to have the same fitness. This model, which stemmed from the assumption (as argued by Motoo Kimura) that the majority of mutations are either extremely deleterious, and therefore rapidly removed from any population, or were effectively neutral (e.g., mutations in noncoding and nonregulatory sequences, synonymous nucleotide substitutions, substitutions of functionally similar amino acids, and the like).

The deterministic dynamics of allele and genotype frequency on a neutral landscape depend only on the mutation rates. In the absence of any directional bias to mutation, the evolutionary process can be modeled as a random walk on an unweighted graph, converging to a uniform equilibrium distribution of genotype frequencies in the limit of infinite time. If finite population effects such as Fisher–Wright genetic drift (i.e., approximation of finite samples from an effectively infinite gamete pool) are introduced, the expected genotype frequencies follow a multinomial distribution on the K genotypes in a population of size N,

$$F(v_1 \ldots v_K) = \binom{N}{v_1 \ldots v_K} f_1^{v_1} \ldots f_K^{v_K}, \qquad (7)$$

where f_i are the initial frequencies of each genotype while v_i are the number of each genotype in the current sample (the sum of all v_i values is N, by definition), implying no expected (average) directional change in the frequency of any particular genotype. This mode of sampling implies that the first and second moments of change in the frequency of any genotype f are given by

$$E[\Delta f] = 0; \quad \mathrm{Var}[\Delta f] = \frac{f(1-f)}{N} \qquad (8)$$

If mutation is nondirectional, then there is no contribution to the first moment term, while the second moment term includes the sample variance in allele frequency and mutation probabilities.

We next consider extensions of these simple models to more realistic scenarios incorporating epistatic interactions.

EPISTASIS AND RUGGED LANDSCAPES

Equation 5a provides an entirely general and dynamically sufficient description of natural selection and point mutation when there are n loci with k alleles. The problem of dimensionality arises from the fact that there are as many fitness coefficients and frequency values (state variables) as there are genotypes. This requires parameterizations

that on the one hand can capture some of the complexity of nonadditivity and epistasis while remaining tractable.

Epistasis involves interaction of allelic effects at multiple loci on phenotype and fitness. From this definition, it follows that one can approximate a fitness function up to arbitrary order using Taylor or Fourier series expansions. Denoting a genotype $\mathbf{z}$ as an n-vector of allelic states $z_1 \ldots z_n$, we can write the fitness in series form:

$$W(\vec{z}) = a_0 + \sum_i a_i S_i + \sum_{i,j} a_{i,j} S_i S_j + \cdots + a_{1,n} S_1 S_2 \cdots S_n. \quad (9)$$

The first term in the series represents the zeroth-order baseline fitness value, it would be the only term in a neutral landscape. The coefficients a_i, $a_{i,j}$, and so on, represent, respectively, the first-order additive contribution of each allele to fitness, the second-order term contributions, and so forth. The polynomial terms S_i associated with alleles z_i in the expansion can have a range in interpretations. In a Taylor expansion, they represent the deviations from mean allelic effects, whereas in a Fourier approximation we have $S_i = \pm 1$. This is a generalization of the model of epistasis familiar to quantitative geneticists, who divide phenotypic variance into additive, additive by additive, and higher-order epistatic component terms.

In heuristic terms, the zeroth-order term generates a flat landscape and the first-order terms give a single-peak landscape, so it follows that higher-order epistatic interactions can generate increasingly more rugged landscapes with multiple local optima, as illustrated in Figure 4.

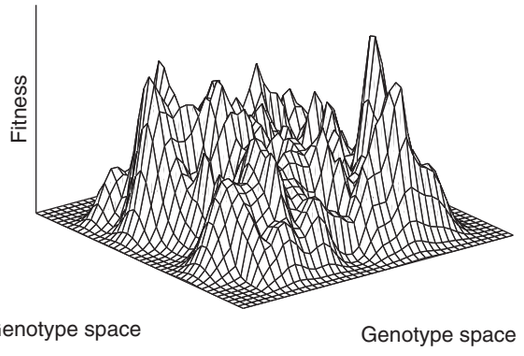


FIGURE 4 A typical rugged adaptive landscape with multiple local optima separated by local minima. Such a landscape has the canonical topography for shifting balance theory and other models, because it requires populations to initially decrease in fitness in order to eventually attain higher fitness.

This iconic image of the rugged fitness landscape has been the source of an extensive research program in evolutionary theory. The specific problem raised by the multipeaked adaptive landscape is the following: the strictly deterministic process of natural selection always favors genotypes with higher fitness. As a result, if the mutation rate is comparatively low (i.e., anything other than single point mutations are negligibly rare), selection will take the population distribution to the nearest peak, behaving as a greedy search algorithm. As a consequence, any population located in the neighborhood of a local optimum would have no way of reaching the global optimum, or for that matter other local optima that have higher fitness. The implication, of course, is that with natural selection and low mutation rates alone, adaptive evolution would rapidly stagnate. One of the principal goals of research in this area is to determine why, in fact, such stagnation at or near local optima does not occur in natural and artificial populations.

Kauffman's *n*–*k* Model

A numerical approach to understanding the dynamics on rugged landscapes was proposed by Stuart Kauffman. He introduced the so-called *n*–*k* model for epistasis, which can be thought of as a special, random-variable case of Equation 9. In the *n*–*k* model, there are n loci contributing fitness effects and k is the order of epistatic interactions. When $k = 1$, a random number (sampled from a uniform or normal distribution) is chosen at each of the n loci, and the genotype fitness is the sum of fitness contributions across loci. With $k = 2$, the fitness contribution of each locus is determined by the sum or mean value of two random variables: its own, and the contribution of one of the randomly chosen $n - 1$ remaining loci. This can be generalized to arbitrary k, where the fitness contribution of each locus is the mean of its own contribution and $k - 1$ chosen at random from the other $n - 1$ sites. It is clear that $k = 1$ corresponds to an additive, single-peak landscape, while higher values of k give a random-variable, kth-order representation of epistasis.

In a series of simulation studies, Kauffman contrasted the outcomes of different search strategies with increasing orders of epistasis. The simplest search strategy involved single point mutations and a random walk that continued for as long as more fit genotypes were accessed in each step. On a $k = 1$ single-peaked landscape, this greedy algorithm always found the global optimum. In

the other limiting case, $k = n - 1$, one has a completely uncorrelated landscape such that the mutational neighbors of a high-fitness genotype are no more likely to have a similar fitness than any random genotype on the lattice. Since random walks terminate once no higher fitness genotypes are accessed, it can be shown that the probability of a random walk of length R is

$$P_R - \prod_{k=0}^{R}\left(1 - \left(\frac{2^k - 1}{2^k}\right)^{n-1}\right), \tag{10}$$

which gives an an expected walk length of

$$E[R] = \mathrm{Log}_2(n - 1) \tag{11}$$

If larger steps are allowed, corresponding to mutations at multiple loci (i.e., $\sim n$ mutations per time step), it can similarly be shown that on an uncorrelated landscape the probability of finding a genotype with higher fitness decreases as a factor of 2 with every step, or, equivalently, that the waiting time to find a more fit genotype doubles with each transition.

Numerical studies have shown that for rugged correlated landscapes ($1 < k << n - 1$), the performance of large transitions is qualitatively similar, entailing a high probability of accessing the neighborhood of a new local optimum initially, but decreasing as a power function with each subsequent step. James Crutchfield and his colleagues have argued that an optimal search strategy for such landscapes is the "Royal Road" algorithm. In this model, the initial search involves mutations at multiple sites, followed by point mutations, with natural selection retaining genotypes along sample paths with higher fitness than the initial state. Essentially, the initial large steps permit the population to access the neighborhoods of local optima with higher fitness, whereas more conservative point mutations allow the population to attain the optimum itself. The model is analogous to a simulated annealing process, in which high-energy particles are allowed to move freely (i.e., passing through regions with lower free energy or high potential energy) initially, followed by a cooling phase in which they settle in the nearest minimum free energy state.

These results raise the question: What biological processes can drive a search strategy that produces large transitions initially before settling into a more conservative (point mutation and selection driven) approach? Interpreted at face value, the large transition steps would involve macromutations, as opposed to the single point mutations during the local hill-climbing phase. Others considering the problem of transitions between local optima have emphasized the importance of genetic drift.

Counteradaptive Evolution and Dimensionality

Genetic drift in finite populations has been proposed as a mechanism that could allow evolving populations to access distant adaptive peaks via low-fitness valleys. Wright's shifting-balance model for evolution was an important elaboration on this theme, as it influenced speculations about the speciation process (see below).

The idea behind these claims is that barring macromutation, the only way to make a transition between peaks along a low-fitness valley is for unfit intermediate genotypes to temporarily increase in frequency through genetic drift. It was shown by Kimura that in sufficiently small populations, genotypes with lower fitness have a nonnegligible probability of increasing in frequency relative to more fit genotypes and even become fixed as a consequence of sampling error in a finite gamete pool.

Although the potential significance of genetic drift cannot be dismissed outright, much of the emphasis on its importance in major evolutionary transitions has been an artifact of our spatial intuition about the shape of adaptive landscapes. In a two- or three-dimensional space, local peaks separated by low-fitness valleys are the de facto norm. However, if the genetic architecture of a trait or fitness value involves higher dimensionality, there is increasing evidence that isolated peaks are a pathological case because in multidimensional genotype or phenotype space there will usually be adaptive "ridges" connecting optima. These results suggest that the topological space that evolution occurs on probably resembles Figure 5 more than the iconographic Figure 4.

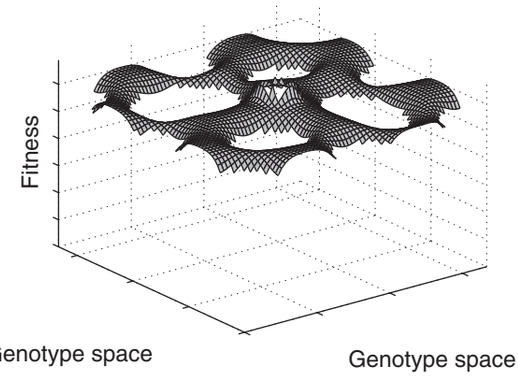


FIGURE 5 An example of a holey landscape, in which all fitness values are either 1.0 or 0. The connected fitness 1.0 genotypes generate a neutral network that does not require counteradaptive evolution in order to access one genotype from another along a mutational path.

One approach that has provided much insight into the topology of multidimensional adaptive landscapes was adopted by Christian Reidys and by Sergey Gavrilets, who made the simplifying assumption of a threshold model of fitness where a genotype in an n-locus, two-allele scenario is either viable or nonviable. If only point mutation is permitted, the sole means of reaching a genotype with a fitness of 1 is via a mutational path along other genotypes with fitness 1. This "Russian Roulette" landscape is constructed by assigning either a fitness of 1 or 0 to each genotype with respective Bernoulli probabilities P, $1 - P$.

As the number of loci (and therefore genotypes) tends to infinity, several patterns in these "holey" landscapes are revealed. First, if $P > 1/2$, the landscape consists of a fully connected neutral network, so that every fitness 1 genotype can be accessed from any other in the landscape through single point mutation. When $P < 1/2$, there are almost always several non-connected neutral networks, so by definition there are several clusters of fit genotypes that are not mutually accessible. However, it was shown by Reidys that when $P > 1/n$, the system is characterized as reaching a percolation threshold, in which the set of fitness 1 genotypes is dominated by a single large component network of viable genotypes, surrounded by comparatively small (including single genotype) viable clusters (see Fig. 6).

In other words, provided that the probability of viability is not very small (i.e., less than the reciprocal of the number of loci), there will always be a dominant "giant" connected component containing $\sim 2^n/n$ genotypes followed by a next largest component of order $\sim n$. The giant component has a radius of size n, meaning that there are viable paths involving mutations at all n loci. These results can be generalized to a landscape where the fitness values can assume values between 0 and 1.

Altogether, these investigations have demonstrated that the problem of accessing optima and moving between local maxima largely disappears with high dimensionality, and that there is no need to invoke genetic drift, macromutation, or other nonadaptive mechanisms to account for the major transitions in evolution.

THE COMPLICATION OF RECOMBINATION

So far, we have only considered mutation as a transmission operator in the evolutionary process. Sexually reproducing organisms produce offspring whose genotypes differ from their own through recombination, which, depending on the number of crossover points and the genetic distance between homologous chromosomes from each parent, can potentially span the entire genotype space.

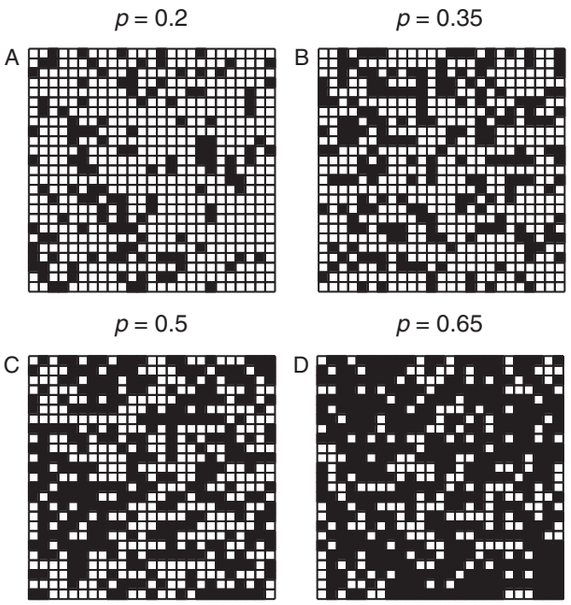


FIGURE 6 Holey landscapes generated by a model in which genotypes have probability P of being viable and $1 - P$ of being inviable. When $P > 0.5$ (i.e., $p = 0.6$ in part (D)), the percolation threshold is attained, so that every viable genotype is a member of a single connected subgraph. For $P < 0.5$, there are multiple subgraph neutral networks, with a giant component dominating when $P > 1/n$ (e.g. $P = 0.45$ in part (C)). As I decreases, the giant component disappears, and the neutral networks become increasingly small and isolated (i.e., $P = 0.15$ and 0.3 in parts (A) and (B), respectively).

Apart from the fact that the genetic (Hamming) distance between recombinants is often greater than that derived through point mutation, there is the additional complication that the transmission-selection dynamics under recombination are inherently quadratic. Assuming that selection occurs before reproduction and recombination in the life cycle, we have

$$x_i(t+1) = \frac{1}{\overline{W}}\left(\sum_{j,k} R_{j,k\to i} w_j w_k X_j x_k\right), \qquad (12)$$

where R is the recombination operator describing the probability that parents with genotypes j, k produce an offspring of type i. Unlike Equation 5, there is no global linearization of the recombination–selection equations.

This nonlinear dynamical system was first analyzed by Richard Lewontin and Ken-Ichi Kojima. The state variables are multilocus haplotypes, with frequencies denoted by $X_{ij\ldots}$ where i, j and other indices denote the allelic state at each of the $1 \ldots n$ loci. Consider the simplest case of two loci and two alleles in a population of haploid, sexually reproducing organisms, such that X_{11}, X_{12}, X_{21}, X_{22} denote the frequencies of genotypes **AB**, **Ab**, **aB**, and **ab**, respectively. If the probability of recombination between the two loci is

given by r, then (for instance) a cross between **AB** and **ab** produces each parental type with frequency $(1 - r)/2$ and each recombinant type (**Ab** and **aB**) with frequency $r/2$.

The selection–recombination equations for the two-locus, two-allele model can be obtained by substituting $(1 - r)$ and r for R, and genotype fitness for w in Equation 10. Except in the limit of nonepistatic selection (i.e., weak selection with approximately additive effects), significant linkage disequilibria across loci are generated, such that in general the frequency $X_{IJ} \neq p_I p_J$ at loci I, J. Consequently, neither the haplotype frequencies nor the mean fitness of the population can be predicted from allele frequencies at each locus. Thus, genetic systems with recombination and epistatic selection offer a clear example of an evolutionary system where allele frequencies are not dynamically sufficient predictors of fitness landscape topology, either in terms of the distribution of genotypes or of mean fitness.

SPECIATION AND SHIFTING BALANCE RECONSIDERED

New insights into the topology of adaptive landscapes have put a new perspective on long-standing notions about speciation and macroevolution. Consider Wright's "Shifting Balance" model for major adaptive transitions. Based on the intuition of a rugged fitness landscape, Wright proposed that the first phase of the process requires genetic drift in a population subdivided into small demes in order to shift genotype frequencies away from a local optimum. This allows genotype frequencies to pass through a low-fitness valley into the neighborhood of a higher adaptive peak. As the genotype frequencies in the population approach a new local optimum, the second phase of shifting balance is largely deterministic, dominated by natural selection taking the subpopulation to the local optimum (i.e., fixation or near fixation of the locally optimal genotype within each subpopulation). The final phase involves a combination of interdemic selection and migration, with populations at higher fitness peaks displacing those occupying lower optima.

Shifting Balance theory was introduced by Wright in large part because it offered a solution to the problem of peak shifts that synthesized selection, mutation, genetic drift, reproductive isolation, and migration into a single process. In spite of its intellectual and aesthetic appeal, a number of recent papers have shown that although it is not impossible for natural populations to follow the shifting balance scenario, they can only do so under rather restrictive parameters. Especially problematic is the first phase. Following Equation 4, the probability of genetic drift traversing a fitness valley is negligible unless the fitness differences are small and the population size is very small. The former condition is problematic because if the fitness differences are minor, then there is no substantial peak shift to speak of, whereas in the latter case there is often little segregating genetic variation for selection (or drift) to act on to begin with.

Perhaps an even stronger critique of the shifting balance theory comes from the fact that in multidimensional genotype space, the peak and valley intuition is misleading as ridges and networks become the dominant topological features (see the section "Epistasis and Rugged Landscapes," above). In the absence of adaptive valleys, there is no need to postulate genetic drift in small populations as a crucial driving force in evolution.

The same issues also arise in population genetic models for speciation. Many contributors to the "New Synthesis," particularly Ernst Mayr, emphasized that speciation necessarily involved a special case of the type of peak shifts associated with phase one of shifting balance. Specifically, Mayr maintained that genetic drift was necessary to traverse fitness valleys and establish reproductive isolation. Consequently, concepts such as "founder flush" speciation and genetic revolutions (sensu Hampton Carson) were introduced as models for allopatric speciation, where drift in very small, peripheral populations lead to drastic changes in the genetic composition of the population, thereby facilitating the peak shift.

Founder flush models, while plausible in theory, suffer from the same problems as phase one of shifting balance—low probability for realistic combinations of parameters, and the fact that they are not necessary if we allow for viable paths between optimal genotypes. As a counterexample, consider the two-locus, two-allele haploid genotypes **AB** and **ab**, both of which have a high fitness. If we assume that recombinants **aB** or **Ab** have low fitness, then there will be strong selection against hybrids and in favor of reproductive isolation. However, this raises the question of how, if mutation rates are low, one attains the **ab** genotype to begin with from an **AB** ancestor, because a single mutation always results in a low fitness intermediate genotype. The simplest solution to this problem was independently proposed by William Bateson, Theodosius Dobzhansky, and Hermann Muller, namely, that one of the intermediate genotypes, for instance **Ab**, was as fit as either **AB** or **ab**, so that there is no need to cross a fitness valley. **Ab** can then mutate to **ab**, which in turn will lead to selection for reproductive isolation with **AB** on account of the unfit **aB** hybrid. An analogous example of an adaptive landscape for two-diploid loci leading to Bateson–Dobzhansky–Muller

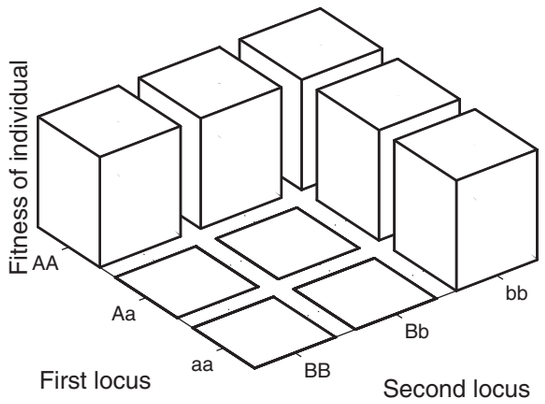


FIGURE 7 An example of a diploid two-locus, two-allele Bateson–Dobzhansky–Muller landscape. Note that the genotypes **AABB** and **aabb** have high fitness, as do **AAbb** and **Aabb**, while the recombinants **aaBB** and **aaBb** are inviable.

(BDM) incompatibilities is shown in Figure 7. As the number of loci increases, the number of possible BDM incompatibilities increase exponentially, as do the number of possible mutational paths connecting viable genotypes. Consequently, contra-adaptive mechanisms such as genetic drift and macromutation need not be invoked as the principal driving forces behind speciation, either.

DISCUSSION

Small changes to parameters determining the order and extent of epistatic interactions can lead to marked changes to the topology of an adaptive landscape. The simplest illustration of this can be seen with an increase in the frequency of viable versus inviable genotypes in the percolation models. However, in order to assess what parameterizations and topologies are biologically relevant, empirical data is necessary.

Measuring the fitness of different genotypes is an extremely difficult task, both in laboratory settings and even more so in nature. It is a nontrivial problem for a combination of reasons—i.e., the statistical difficulties of inferring fitness components from competition experiments, the problem of genotyping (becoming less of an issue today on account of high-throughput sequencing technology), and above all the issue of controlling for extraneous environmental variables that confound survival or reproductive advantage in competition experiments.

Consequently, some of the most robust studies of fitness landscape structure don't involve whole organisms, and instead focus on in vitro studies of the catalytic performance or the thermodynamic and/or structural stability of macromolecules such as RNA and proteins. In the case of RNA, the biophysics of folding are sufficiently well understood to predict fitness (i.e., mean square distance to an optimal structure, or to a free energy minimum). Walter Fontana and colleagues analyzed the topology of RNA landscapes using a random walk autocorrelation measure ρ, which provides a measure of the landscape ruggedness (i.e., a high autocorrelation score suggests a relatively smooth fitness surface with few peaks, while a low score implies a rugged landscape where a single nucleotide substitution can lead to radical changes in secondary structure). The authors compared the autocorrelation scores to those obtained on landscapes using Kauffman's n–k model and found that the RNA landscape was consistent with a model where $k \approx 7$, for $n \sim 100$.

Another approach to the problem of inferring the shape of adaptive landscapes is indirect, involving comparative sequence analysis across taxa. For example, comparing human protein sequences to those of other mammals, Alexey Kondrashov and his collaborators found that amino acid substitutions at sites associated with genetic diseases in humans are viable in other mammals such as the mouse. This suggests that epistatic interactions across sites are crucial in maintaining (or restoring) function. Since the mammalian sequences in the sample shared common ancestors with humans ~100 million years (and many speciation events) ago, this suggests a path of viable point mutations connecting the orthologs and provides evidences for BDM-type incompatibilities discussed in the previous section.

These examples are cited as two case studies that are representative of a growing literature among researchers using similar strategies—i.e., attempts to predict fitness from first principles based on macromolecular biophysics, and comparative approaches of homologous sequences across different species. In the future, as our knowledge of gene regulatory networks and quantitative trait loci in model organisms improves, it should be possible to make predictions about the structure of adaptive landscapes for more complicated epistatic model systems.

In addition to these empirical studies, there remains the important task of characterizing the evolutionary dynamics on multidimensional landscapes as our understanding of their structure improves. As was the case at the field's inception, there remains the daunting problem of high dimensionality in mutation–selection on landscapes in multilocus epistatic systems. The problem is even more pronounced for recombination–selection systems because of their nonlinearity. Most recent studies have either made use of approximations involving the first few moments of quantitative trait distributions or assumptions about the symmetry properties of fitness functions to reduce the

numbers of variables and parameters necessary to characterize evolution. These methods will be increasingly valuable as more elaborate, data-driven models of epistatic interactions and pleiotropic effects are generated.

SEE ALSO THE FOLLOWING ARTICLES

Adaptive Dynamics / Evolutionary Computation / Mutation, Selection, and Genetic Drift / Quantitative Genetics

FURTHER READING

Coyne, J. A., N. H. Barton, and M. Turelli. 1997. A critique of Wright's shifting balance theory of evolution. *Evolution* 51: 643–671.

Eigen, M., J. MacCaskill and P. Schuster. 1989. The molecular quasispecies. *Advances in Chemical Physics* 75: 149–63.

Fontana, W., P. F. Stadler, P. Tarazona, E. O. Weinberger, and P. Schuster. 1993. RNA folding and combinatory landscapes. *Physical Review E.* 47: 2083–2099.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton, NJ: Princeton University Press.

Kauffman, S. A. 1993. *The origins of order*. Oxford: Oxford University Press.

Kimura, M. 1980. *The neutral theory of molecular evolution.* Cambridge, UK: Cambridge University Press.

Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky–Muller incompatibilities in protein evolution. *Proceedings of the National Academy of Sciences* 99: 14878–14883.

Reidys, C. M., and P. F. Stadler. 2001. Neutrality in fitness landscapes. *Applied Mathematics and Computation* 117: 321–350.

van Nimwegen, E., J. P. Crutchfield, and M. Mitchell. 1999. Statistical dynamics of the Royal Road genetic algorithm. *Theoretical Computer Science* 221: 41–102.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selectin in evolution. In D. F. Jones, *Proceedings of the 6th International Congress on Genetics,* Vol. 1. pp. 356–366. Austin, TX.

AGE STRUCTURE

TIM BENTON

University of Leeds, United Kingdom

An individual's age is often a strong determinant of its life history. Age can be used to stratify individuals into different age classes, with the rationale that individuals within the same age class will have similar life histories and therefore similar demographic rates. Such a population is termed age structured. Modeling age structure in discrete time is commonly undertaken with matrix models; in continuous time, with partial or delay differential equations. Age- (and stage-) structured models have been influential in many areas of evolutionary, population, and conservation ecology.

WHAT DOES AGE DO?

Predicting population dynamics with any precision requires some sort of population model. The simplest conceptualization is an "unstructured" model, where the future population size is some function of the current (or past) total population size; i.e., $N_{t+1} = f(N_t)$. This formulation implies that variation between individuals in age does not matter. This may be because there is an assumption that on the scale of the model individual differences are averaged out or age may not be relevant. For example, modeling populations of annual plants based on annual census data does not need to take into account the age of the plants. However, in other circumstances, age differences between individuals are crucial. All individuals go through a life cycle: they are born (sprout or hatch) small, grow as they develop, mature and reproduce as prime-aged adults, and then decline during senescence in reproductive rates and survival. At some spatial or temporal scale, a population's demographic rates will always depend on its age profile or structure, but it is sometimes possible to ignore this—for the example of annual plants, age structure can be ignored at the annual temporal scale, but it would become important when looking at their dynamics at a monthly scale.

To illustrate the importance of age structure, two human populations are contrasted in Figure 1: the population of India has a greater proportion of young people than the UK (India has 43% of its population between 15 and 39, whereas the UK has 33%), and far fewer older people (8% over 60 versus 22%). These differences in age structure are reflected in the populations' demographic rates (i.e., per capita reproduction and survival) and ultimately cause differences in the population growth rates: India's population is growing almost twice that of the UK.

Age is not the only factor that differentiates between the N individuals in a population, and therefore it is not the only factor that influences demographic rates and creates population structure. Sex ratio may be important, and for some systems, there may also be differential performance between individuals depending on their body size, quality, or condition; so population structure can arise not only due to age, but it can also be size structured, or even size-and-age structured (see below in the discussion of Integral Project Models below). In many invertebrates, adults and larvae may have very different demographic rates, and so populations may be structured according to the developmental stage they are in. In many invertebrate species, developmental duration can be very plastic depending on food availability and temperature, so knowing an individual is a certain age will tell you less about their vital rates than knowing they are an adult or a pupa.

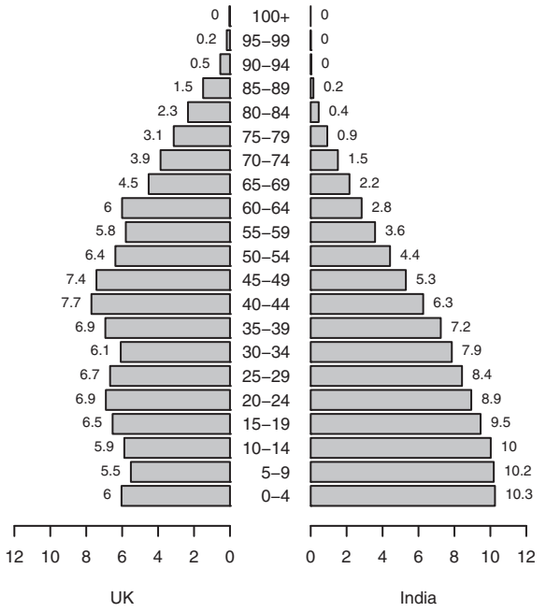


FIGURE 1 Population pyramids for India and the UK, 2009. Each bar shows the proportion of the total population in an age group. The Indian population has a much greater proportion of young people and people of reproductive age. Data from the U.S. Census Bureau, International Database (http://www.census.gov/ipc/www/idb/index.php). The population growth rate for the UK is 0.7 (though some of this is driven by immigration); for India it is 1.34 (data from http://data.worldbank.org/).

The importance of incorporating the life cycle into population models (whether as stage or age) is that it explicitly incorporates history: individuals born at time t cannot reproduce until they have developed and passed into the correct stage, age, or size class. This "history in the life history" is particularly important when modeling the way populations respond to changes in the environment (whether the biotic environment, such as density dependent processes, or the abiotic environment, such as changes in the climate), as the population's response to a change may lag behind the change. The characteristics of the lagged response will be determined by the population's structure and may therefore change with time. For example, if food suddenly becomes available in a population with many adults, there could be a quick increase in reproduction leading to a fast response. Conversely, a population with many juveniles would respond more slowly, as the juveniles would have to grow to adulthood in order to reproduce and therefore boost population growth rate. Lags in dynamical systems also have the property of introducing instability, and so lags due to generation time are one of the common causes behind population cycles. In general, structured models may have quite different dynamical properties from unstructured models with the same average demographic rates.

Age-structured models incorporate age structure into the model and describe the dynamics of the whole population as the sum of the subpopulations in the different age classes. Age-structured models can naturally be formulated in discrete or continuous time. Discrete-time models are often appropriate for species that reproduce annually and may be censused each year (which includes populations of many vertebrates). Such discrete-time models encapsulate an equation that gives transition rates (i.e., survival) from one age class to the next, and also reproductive rates for each age class. The reproductive rates are strictly called fertilities, as they incorporate both reproduction and the survival of the offspring until the next time step. The classical, discrete-time, age-structured model is called a Leslie matrix model (see the section below) and is an example of the broader family of matrix models. In continuous time, the McKendrick–von Foerster equation is an analogue of the Leslie matrix, but because age structure introduces lags into the dynamics, another natural way to deal with age structure is via delay-differential equations (DDEs).

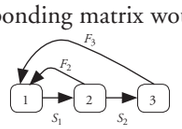
THE CLASSIC AGE-STRUCTURED LESLIE MATRIX MODEL

The classic age-structured, discrete-time, matrix model is often called a Leslie matrix model, after P. H. Leslie (a colleague of Charles Elton at the Bureau of Animal Populations at Oxford), who was an early pioneer of their use in ecology. The Leslie matrix contrasts with the Lefkovitch matrix, which is a stage-structured matrix model.

At time t, the population is represented by a vector n_t with each element n_i corresponding to the number of individuals in age class, i ($1 \geq i \geq \max(age)$); so the total population size $N = \sum_1^{\max(age)} n_i$. The population at time $t + 1$ is projected using a recurrence equation,

$$\mathbf{n}_{t+1} = \mathbf{A}\mathbf{n}_t, \tag{1}$$

where $\mathbf{A}$ is a square matrix. For a simple three-age class model, with individuals growing to age 3 (then dying), reproducing at ages 2 and 3 (a life cycle similar to many small passerine birds perhaps), the life-cycle diagram and corresponding matrix would be as follows:



$$\mathbf{A} = \begin{pmatrix} 0 & F_2 & F_3 \\ S_1 & 0 & 0 \\ 0 & S_2 & 0 \end{pmatrix}. \tag{2}$$

F_i is the fertility, the number of offspring that are born to an adult in age class i and which survive to the next census, S_i is the proportion of individuals in age

class i that survive to age class $i + 1$ at the next census. An age-structured matrix is always one that has nonzero elements only in the top row and in the subdiagonal. In the classic model, the values of F_i and S_i are constant. If Equation 1 is iterated forward sufficiently, the influence of any arbitrary starting vector **n** disappears and the population grows at a constant rate, λ, and the age structure (i.e., the proportion of individuals in each age class) becomes constant: this is the stable age distribution (see Fig. 2). This property that initial conditions make no impact on the long-run (asymptotic) behavior is known as ergodicity. Rather than iterate Equation 1 to find the asymptotic properties of the matrix, they are available as the solution of the characteristic equation of Equation 1 (i.e., $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, where "det" is the determinate, and **I** is the identity matrix). The solution gives a vector of eigenvalues (of which there are as many as age classes), but the largest, or principal, eigenvalue is λ, the population growth rate. When the eigenvalues are known, the equation $\mathbf{Aw} = \lambda\mathbf{w}$ can then be solved. The vector **w** is the right eigenvector (as it sits on the right of the matrix **A** in the equation). The left eigenvector, **v**, is found from $\mathbf{v}'\mathbf{A} = \lambda\mathbf{v}'$. The right eigenvector corresponds to the stable age distribution, and the left eigenvector corresponds to the reproductive value of each age class (i.e., the relative contribution to population growth from each age class and which typically increases up to maturity and then declines). Clearly from Equation 1 if the matrix **A** changes, the characteristic equation changes and so do the eigenvalues and eigenvectors as well as, therefore, their biological interpretations: population growth, stable age distribution, and age-related reproductive values. In fact, some of the appeal of matrix modeling arises from the fact you can easily analyze the way λ will change with changes in the matrix. As each element of **A** (the a_{ij}) is varied, λ will change and the rate of change in λ with respect to a change in a_{ij} is given by

$$\frac{\partial\lambda}{\partial a_{ij}} = \frac{v_i w_j}{\langle \mathbf{w}, \mathbf{v} \rangle}. \tag{3}$$

Here, v_i is the ith element of vector **v**, and $\langle \mathbf{w}, \mathbf{v} \rangle$ represents the scalar product of the two vectors. The quantity $\partial\lambda \backslash \partial a_{ij}$ is called the sensitivity of λ and is the absolute rate of change in λ with a change in each matrix element. The elasticity is the sensitivity rescaled to be the proportional change in λ with a change in the matrix element. The mathematical formulation is the same whether the matrix model is age, sex, size, or stage structured. Figure 2 gives an example of an arbitrary Leslie matrix iterated over a number of time steps, plus the simple analytical output.

The classical Leslie matrix model (and its related stage-structured formulations) has been hugely influential in both population management and evolutionary ecology. Much of the popularity comes from the work of Hal Caswell (see Further Reading for Caswell's iconic book). Parameterizing a matrix model from

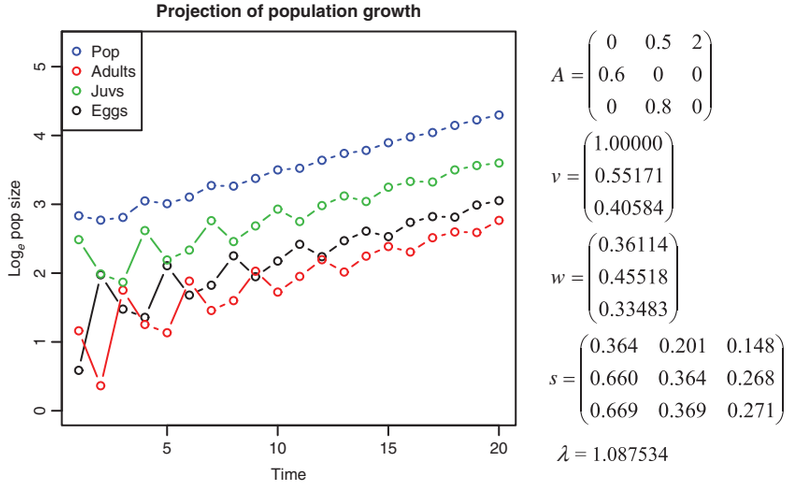


FIGURE 2 Example of an age-structured model. Matrix **A** is multiplied by an arbitrary initial vector (**n** = (2, 4, 7)′), and this is iterated for 20 time steps. The population growth settles down to a constant rate (linear gradient on a log scale), λ = 1.088 (on a log scale that is a gradient of 0.084); with the age distribution being given by **w**, the right eigenvector (i.e., 36% of the population in age class 1, 46% in 2, and 33% in 3); the reproductive value being given by **v**; and the matrix of sensitivities by **s**. Thus, λ is most sensitive to changes in the survival of individuals from age 1 to 2 (s_{21} = 0.66). This suggests, all being equal, that a management intervention that increases S_1 would have the largest impact on λ compared to changing any other vital rate. Note, the largest sensitivity a_{31} = 0.669 is for an impossible transition (moving from age 1 to age 3 without passing through age 2).

census data may often be quite striaghtforward, and in common software (like MATLAB, Mathematica, or R) the eigenanalysis and calculation of sensitivities may only take a couple of lines of code. Hence, matrix models in their simplest form are characteristically uncomplicated; one can estimate population growth rate and how, by altering different vital rates, it may respond most productively to interventions. Typically, in Leslie models λ is most sensitive to changes in survival rates rather than fertilities. Sensitivities have been used, for example, in turtle conservation to switch some effort from protecting eggs on beaches to preventing mortality of adults via fishing. Furthermore, fitness can be identified as the population growth rate of a gene's lineage, and so a matrix with a larger λ will invade a population whose λ is smaller. This means that the sensitivity of λ to changes of the a_{ij} measures the selection pressure on the matrix element. If a small change in the element causes a large change in fitness, it will be strongly selected; if, conversely, a small change in the element has no effect on λ, it will not be selected. Therefore, this simple modeling framework has been influential in studies of evolutionary ecology as well as population and conservation ecology.

EXTENSIONS OF THE LESLIE MODEL

The classic Leslie matrix model is time invariant, in that the vital rates (the a_{ij}) are assumed to be constant. Such a formulation is useful for population projection (i.e., asking "What would happen *if* . . . ?" questions) but are less good for population prediction (asking "What *will* happen?"). This is because in any real-world situation demographic rates vary over time. Time-varying rates occur because the environment varies (and varies over multiple time scales: daily, seasonally, annually, decadally, and also may change permanently due to a range of factors, such as habitat destruction, climate change, and so on). Also, any realistic system will be regulated by density dependence around a dynamical attractor; any small perturbation from the attractor will be followed by a return that is governed by density-dependent processes. Time-varying vital rates at a population level correspond with changes in individuals' life histories. Many of these changes will reflect phenotypically plastic changes in the life history that results from changes in resources to different traits (for example, a severe winter may reduce population density, which, in turn, reduces resource competition in the spring, giving greater access to resources and allowing an increase in survival, growth, or reproduction).

The simplest way to build in variability in the vital rates is conceptually to assume that, for each time step, the vital rates are drawn from some probability distribution, with the mean being the observed parameter. A number of studies have conducted simulations that have resampled matrices at each time step and iterated the population dynamics over a long time period to give population trajectories in stochastically varying environments. From the observed dynamics, one can estimate both the stochastic population growth rate (λ_s) and the stochastic sensitivities by numerical differentiation. Conceptually, simulating population trajectories with time-varying vital rates underpins population viability analysis (PVA), in that PVA estimates the distribution of population sizes for a stochastic model at some point in the future; so PVA is an important application of stochastic structured modeling. In addition to simulation studies, considerable efforts have been made in recent years to develop stochastic matrix theory, and there is now a range of formulae (backed up by freely available computer code) that allows approximate analytical estimation of the standard quantities. For example, Shripad Tuljapukar published a formula in 1990 for estimating the stochastic population growth rate. The simplified version is given here:

$$\log \lambda_s \approx \log \lambda - \frac{1}{2}\left(\frac{\tau^2}{\lambda^2}\right). \qquad (4)$$

Log λ_s is therefore approximated by the log of the population growth rate of the mean matrix (λ) minus a quantity that is a function of τ^2, which is itself a summation across every possible pair of matrix elements of the covariances between those elements, multiplied by the sensitivity of λ for each of those elements. τ^2 is therefore the approximate variance of the population growth rate that is caused by variation in the a_{ij}. From this formula a general point emerges: temporal variation in the vital rates will typically lead to a reduction in population growth rate (or fitness). Those rates with a high sensitivity will disproportionately decrease growth rate if they vary, so selection should tend to minimize variation in those rates. However, this general conclusion depends on the way that the vital rates covary. If there is strong negative covariation between rates (e.g., a good year for fecundity is a bad year for survival), the net effect can be that variation in some traits may be selected for, as variation in one trait can counteract the variation in another.

The stochastic demography above assumes that, although they may covary, traits vary independently of each other: changes in one vital rate variance can occur independently of changes in any other. This is clearly not

necessarily the case. Life history theory is built around resource allocation tradeoffs (where to increase one vital rate necessarily reduces another); these can readily be incorporated into age-structured models. Likewise, density dependence will create covariation in vital rates: if the density goes down, per capita access to resources goes up, and so individuals may be able to increase more than one vital rate simultaneously. In theory, it is quite straightforward to introduce density dependence into vital rates, but in practice this may be fraught with methodological difficulties. For example, there is rarely sufficient high-quality information about local density changes coupled to changes in vital rates, so knowledge is incomplete. In addition, the estimation of the temporal or spatial scale of density measurements is complex. The performance of a juvenile may be related to the density in its current neighborhood, but it may also be related to the density in its mother's neighborhood when she was allocating her resources to him. So identification of the "correct" density relationship with one or more vital rates may be problematic. Nonetheless, density-dependent versions of structured models have been investigated, and one interesting result emerges in the time-varying case: the sensitivities of λ_s predict the effects of changes on fitness but not on population size. An increase in a vital rate can cause a decrease in population size (for example, an increase in fecundity, all things being equal, will increase fitness, but if it increases the population density of the juveniles so much it reduces their survival, it may decrease the total population size). Density-dependent models have been explored numerically, but recently Caswell has developed considerable theory for the estimation of elasticities of nonlinear models and transient dynamics (see Further Reading).

THE INTEGRAL PROJECTION MODEL

The Leslie matrix model structures the population by age and ignores all other information. In reality, within each discrete age class (or stage class in the stage-structured case) there will inevitably be variation between individuals in other important characteristics. For example, adults of the same age may differ continuously in size or quality, both of which may cause predictable change in their survival and reproductive rates. Such variation in characters and traits can be incorporated into an age-and-stage–based model (e.g., adults of age i can be subcategorized as large and small), but anything more than a few categories quickly becomes unmanageable as the matrix increases dimension. Integral projection models (IPMs) are a way to combine trait, or character, variation into a structured model. IPMs are constructed around functions that relate variation in characters or traits to demographic rates. These functions can vary with age (or stage) and may vary over time due to environmental variation or changes in population density. In general, there are only four main relationships needed to fully describe demographic rates in a population: the associations between the structuring variables (e.g., age and size) and (a) survival, (b) development among those that survive, (c) fertility, and (d) the variation in offspring size that results from parents with the observed values of age and size.

An important concept in IPMs is that the population structure comprises a set of components, which are a combination of discrete classes (such as age classes) and continuous domains (such as size functions within each discrete class). The vector of population sizes, $\mathbf{n}$, is replaced by a set of distribution functions $n_i(x, t)$, where $n_i(x, t)dx$ is the number of individuals in age class i at time t with their state variable (such as size) within the interval $[x, x + dx]$. The projection matrix $\mathbf{A}$ is replaced with a set of kernels, $K_i(y, x)$, that represent, within age i, the survival and growth of state x individuals to state y, or the production of state y offspring by state x parents. The dynamics of the $n_i(x, t)$ are given by a set of coupled integral equations:

$$n_i(y, t + 1) = \sum_{j=1}^{C} \int_{\Omega_j} K_{ij}(y, x) n_j(x, t) dx, \qquad (5)$$

where C is the total number of components (e.g., age classes or age x stage classes), Ω is a closed interval characterizing the size domain, and $1 \geq i, j \geq N$. This model is density independent, but it can easily be made density dependent by making the kernels responsive to a measure of population density (i.e., $K(y, x, d)$), or by making the kernels time variant, or even both (i.e., $K(y, x, d, t)$). For a set of assumptions, there will be stable population growth, and equivalent calculations can be made to calculate population growth rate, stable age distributions, reproductive values, and sensitivities.

So, it is possible to build models based on trait-demography associations that describe age-specific functions linking the trait to survival, fertility, development of the trait in individuals that survive, and reproductive allocation to offspring. These models can be used to describe a range of metrics, including population growth, selection differentials, descriptors of the life history (such as generation time), and even estimates of heritabilities of traits. This approach has recently been illustrated by Tim Coulson and colleagues (see Further Reading). This indicates that the coupled life history, phenotypic, population, and

evolutionary dynamics can be extremely complex. Perturbing a vital rate, or a trait, at one age can have effects later on in the life history and impact upon later generations, with consequences both ecological (i.e., changes in population structure and growth rate) and evolutionary (changes in selection and the response to it). In Soay sheep, for example, an increase in reproduction by young adults leads to a decrease in mean body size and also a decrease in the selection for larger body size. Understanding how populations respond to environmental change (perturbations) in ecological and evolutionary terms therefore requires modeling the associations between traits, ages, and population structure, and, without a comprehensive model, accurate prediction of how a system will respond will not be possible. The complexity of causation of dynamics is increasingly being recognized by both life historians and dynamicists, and it is backed up from our in-depth understanding of a range of well-studied model systems. The IPM framework therefore provides a detailed way of modeling this complexity and therefore fully understanding, and predicting, the system responses to environmental change. The simplicity of the classic Leslie matrix is a double-edged sword in that it is very easy to implement but also only really useful for projection, not prediction.

This entry has focused on discrete time models, partly because we naturally discretize age (incrementing it by 1 unit annually on our birthday), and partly because our understanding of the population dynamics of many systems is based on annual censuses, with the biological systems having set, discrete reproductive seasons (rather than continuous reproduction throughout the year). It is, of course, perfectly possible to model age-structured populations using a continuous time framework. McKendrick introduced a continuous time model in 1926 that describes the dynamics of the age distribution using a model now referred to as the McKendrick–von Foerster equation. If $n(t, a)$ represents the age distribution at time t, the McKendrick–von Foerster equation is

$$\frac{\partial n(t,a)}{\partial t} + \frac{\partial n(t,a)}{\partial a} = -\mu(t,a)n(t,a), \qquad (6)$$

where $\mu(t, a)$ is the specific death rate for age a and time t. The McKendrick–von Foerster equation expresses the dynamics of the population with mortality and aging, but has no reproduction. To complete the formulation, one needs a boundary condition for the birthrate:

$$n(t,0) = \int m(a,t)n(a,t)da,$$

where $m(a, t)$ is the birthrate of individuals at time t from age a mothers.

This model has been extended such that it can be used to describe dynamics in fluctuating and density-dependent environments. It can also be applied, much like the IPM, when more than age (e.g., age and size) has a strong impact on an individual's vital rates. Unlike matrix models, these general physiologically structured models are complex to analyze because they are based on partial differential equations. There is an intermediate type of model, called the escalator box car, that can span the continuum from a classic matrix model to a continuous time version of the McKendrick–von Foerster equation; this has the benefit of being able to combine the computational ease of matrix models with the flexibility of continuous time models.

It is also possible to model in continuous time an age-structured population using coupled delay differential equations, but logically these are perhaps best thought of as stage-structured models and are well described in that entry.

INTRODUCTION TO THE LITERATURE

Age-structured theory is advancing apace. The definitive text for matrix models remains Caswell (2001), which also includes an introduction to the problem of parameter estimation for the models and some of the basics of stochastic and density-dependent formulations. A general review about the importance of individual variation, and hence the need for structured models, is given by Benton et al. (2006). The classic reference for stochastic matrix analysis is found in Tuljapurkar (1990). A good review of stochastic matrix models and their use in applied biology can be found in Fieberg and Ellner (2001). Caswell has developed a considerable body of theory recently on sensitivity analysis of nonlinear models (Caswell 2008) and models with transient dynamics following a perturbation (Caswell 2007). A worked example, using numerical approaches, of a density dependent matrix model applied to the well studied flour beetle, *Tribolium,* is given in Grant and Benton (2003). Integral projection models are reviewed in Ellner and Rees (2006), which both discusses the general theory and gives a thoroughly explored example for a thistle. Well-studied ungulate systems have been particularly instrumental in highlighting applications of the theory discussed above. For example, Coulson et al. (2003), on red deer, apply Van Tienderan's extension of sensitivity analysis and link selection gradients between a phenotypic trait and multiple fitness components with the effects of these fitness components on the population growth rate (i.e., mean absolute fitness). One of the most synthetic papers of recent years is Coulson et al. (2010),

which uses the Soay sheep data from St. Kilda, Scotland, and investigates the joint eco-evolutionary dynamics by using a trait (size)-based, age-structured model to estimate heritabilities, selection differentials, and life-history descriptors in addition to population growth parameters.

SEE ALSO THE FOLLOWING ENTRIES

Delay Differential Equations / Matrix Models / Population Ecology / Partial Differential Equations / Population Viability Analysis / Stage Structure / Stochasticity

FURTHER READING

Benton, T. G., S. J. Plaistow, and T. N. Coulson. 2006. Complex population dynamics and complex causation: devils, details and demography. *Proceedings of the Royal Society B: Biological Sciences* 273: 1173–1181.

Caswell, H., 2001. Matrix Population Models—construction, analysis, and interpretation. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.

Caswell, H. 2007. Sensitivity analysis of transient population dynamics. *Ecology Letters* 10: 1–15.

Caswell, H. 2008. Perturbation analysis of nonlinear matrix population models. *Demographic Research* 18: 59–113.

Coulson, T., L. E. B. Kruuk, G. Tavecchia, J. M. Pemberton, and T. H. Clutton-Brock, 2003. Estimating selection on neonatal traits in red deer using elasticity path analysis. *Evolution* 57: 2879–2892.

Coulson, T., S. Tuljapurkar, and D. Z. Childs. 2010. Using evolutionary demography to link life history theory, quantitative genetics and population ecology. *Journal of Animal Ecology* 79: 1226–1240.

Ellner, S. P., Rees, M., 2006. Integral projection models for species with complex demography. *American Naturalist* 167, 410–428.

Fieberg, J., Ellner, S. P., 2001. Stochastic matrix models for conservation and management: a comparative review of methods. *Ecology Letters* 4244–266.

Grant, A., and T. G. Benton. 2003. Density-dependent populations require density-dependent elasticity analysis: An illustration using the LPA model of *Tribolium*. *Journal of Animal Ecology* 72: 94–105.

Tuljapurkar, S. 1990. *Population dynamics in variable environments*. New York: Springer-Verlag.

ALLEE EFFECTS

CAZ M. TAYLOR

Tulane University, New Orleans, Louisiana

In the 1930s, Warder C. Allee demonstrated that cooperation between individuals can be as important as competition. Individuals can suffer decreased fitness when they are in groups in which there are too few conspecifics or in which the individuals are too diffuse. This simple observation, now called the Allee effect, has far-reaching consequences, affecting population dynamics, extinction rates, and invasion rates, as well as evolution. The most dramatic dynamical consequence of an Allee effect happens in cases when density levels lower than a threshold drive the per capita growth rate below replacement rate and the population begins to decline, eventually to extinction.

CONCEPT AND DEFINITIONS

The formal definition of a component Allee effect is the existence of a positive relationship between a component of fitness and population size or population density. This means that survival or reproduction is decreased when density of conspecifics is low. A component Allee effect can lead to a demographic Allee effect when the overall per capita growth rate is also decreased at low densities and shows a positive relationship with population density or size. A demographic Allee effect can be either strong or weak. When a population experiences a weak demographic Allee effect, the per capita growth rate is lowered at low densities but never drops below replacement rate so that the population does not actually decline but increases more slowly. When a population experiences a strong demographic Allee effect, the per capita growth rate drops below replacement rate when the population density is lower than a threshold value (called the Allee threshold). Below the Allee threshold, a population subject to a strong Allee effect declines to extinction.

Allee effects are also known as positive density dependence, in contrast to (negative) density dependence, the negative relationship between density and fitness due to competition between conspecifics. At the demographic level, competition is a negative feedback (or compensatory) because the growth rate drops as the population increases, usually stabilizing the population at a carrying capacity. Allee effects, by contrast, are a positive feedback and are called depensation or depensatory dynamics, especially in the fisheries literature, as opposed to compensation or compensatory dynamics.

MECHANISMS

There are multiple mechanisms by which an individual can suffer from insufficient numbers or density of conspecifics. Broadly speaking, for an Allee effect to occur, there must be a positive relationship between a component of fitness and population density at low densities. Fitness is the contribution an individual makes to future populations and has many components that can be classified roughly into those that affect survival and those that affect reproduction. Thus any mechanism that lowers any component of survival or reproductive

success at low densities is a mechanism of a component Allee effect. The two most commonly cited general causes of Allee effects are the difficulty of finding a mate (or viable gametes) at low densities in sexually reproducing species and the higher likelihood of being depredated at low densities. Other cooperative behaviors can also cause Allee effects when there are too few conspecifics to convey the benefit of the cooperative behavior. Some Allee effects are caused by the behavior of other, interacting species like pollinators or predators, and sometimes Allee effects are attributable to genetic causes.

Mate Finding

Perhaps the most commonly cited mechanisms of Allee effects are based on the idea that, at low population density, sexually reproducing species may have difficulty finding a (or enough) suitable, receptive mates (or gametes) and reproductive output is decreased. This occurs in both mobile organisms that actively seek mates and in sessile organisms that rely on transport of gametes through surrounding air or water. When the transport is largely passive, a cloud of sperm or pollen diffuses and becomes more dilute with increasing distance from the source organism. Consequently, in sparse populations, widely separated individuals suffer decreased fertilization.

Predation

Another very common mechanism of an Allee effect is the predation dilution effect caused by aggregation in order to reduce the per capita probability of being depredated. Further reduction in per capita mortality from predation, and therefore more benefit from being in large groups, can be achieved in species that form such large aggregations that predators are swamped or satiated. Aggregation can occur both in space and in time and potentially includes phenomena such as masting, where seed production of all individual plants in a population is synchronized. Additionally, terrestrial predators may attack the edge of an animal group (e.g., a herd of ungulates). There will tend to be proportionally fewer individuals on the edge of a large group than a small group, so per capita mortality is reduced in a larger group. Furthermore, animals in large groups may be able to devote less time to vigilance than in a small group but achieve the same or better protection by relying on their conspecifics to raise an alarm if a predator approaches. Time saved in vigilance can be spent foraging or on other activities that improve fitness. Animals may also engage in other cooperative behaviors that are designed to confuse predators or otherwise reduce predation.

Cooperative Behaviors

Allee effects resulting from decreased predation risk are caused by aggregative or more active cooperative behaviors on the part of the species. Other types of cooperative behaviors can also result in Allee effects. Examples include cooperative foraging, in which animals hunt in packs and large groups and are therefore more efficient at catching or finding prey; reproductive facilitation, in which individuals are stimulated to reproduce by the presence of reproductive conspecifics; cooperative breeding, in which large groups improve juvenile survival; female choice, where females may choose not to mate at all if the selection of males is too small; and environmental conditioning, in which species are able, when in large groups, to modify or condition their environment via several different mechanisms (for example, thermoregulation) in such as way as to improve survival or reproduction.

Behaviors of Interacting Species

Sometimes an Allee effect is caused by the behavior of an interacting species rather than by the species experiencing the Allee effect. A species that benefits or provides a service to another species will cause an Allee effect if it prefers to service individuals in large groups. The most common example, and one that could be classified under the mate-finding section, is pollinator limitation. Some pollinators are less likely to visit small groups of flowering plants than to visit larger groups. When this is the case, animal-pollinated plants in smaller groups may experience lower pollination and seed production than those in larger groups.

In the reverse mechanism, a harmful species like a predator that preferentially preys on individuals in small groups will cause an Allee effect in the prey. The clearest example of this type of Allee effect occurs when humans act as the predator either hunting or collecting species. It has been shown that humans put a high value on rarity for its own sake and species that are rare or in danger of extinction often experience increased hunting or collection pressure or simply disturbance from people wanting to own a specimen or get a glimpse of a disappearing species.

Genetic Mechanisms

In small populations, and particularly in suddenly reduced populations, genetic diversity tends to be lower than in large populations. Low genetic diversity often

leads to a decrease in individual fitness. These two facts put together lead to genetic Allee effects. Mechanisms that lead to low genetic diversity in small populations are genetic drift, the sampling effect, and increased inbreeding. The sampling effect is a loss of allelic richness due to a relatively abrupt reduction in the number of individuals in the population, and genetic drift is the continuing loss of alleles in small populations because chance mating events mean that some alleles drop out in each generation. Increased inbreeding leads to increased homozygosity. A drop in allelic richness allows the fixation of mildly deleterious alleles and a drop in beneficial alleles and a resulting decline in individual fitness. Increased homozygosity reduces fitness because it allows the increased expression of deleterious alleles and also when homozygotes are less fit than heterozygotes.

POPULATION MODELS AND DYNAMICS

To include an Allee effect in a mathematical population model, there are two basic approaches. One is to model the demographic Allee effect directly by creating a phenomenological model, which means creating a population model in which some part of the per capita growth rate has a positive slope when plotted against population density (Fig. 1). The other approach is to separate the different components of fitness and include each separately in the model and then to include the Allee effect only in the component of fitness in which it is know to be present. Component Allee effects do not always lead to demographic Allee effects, and this can be shown by this second approach. On the other hand, multiple component Allee effects are possible in a single population and

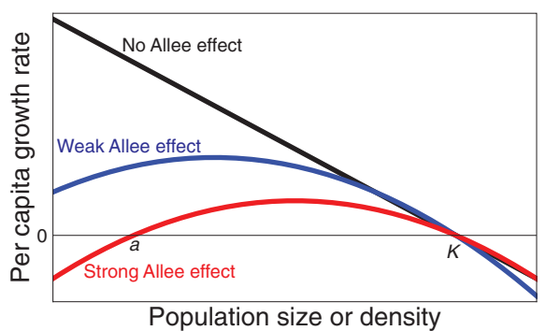


FIGURE 1 *Per capita* growth rate of a population. With no Allee effect (black line), the *per capita* growth rate declines with population size or density. In populations with both any kind of Allee effect (red and blue lines), there is a positive relationship with density at low densities. A strong Allee effect (red line) creates a threshold (a) below which the *per capita* growth rate is negative and the population will become extinct. A weak Allee effect (blue line) has no threshold only a slower growth rate at low densities than at higher densities.

can interact with one another, giving interesting population dynamics.

Models of Demographic Allee Effects

A very general population model in continuous time can be written

$$\frac{dn}{dt} = nf(n),$$

where n is population density and $f(n)$ is the density-dependent per capita growth rate. Any functional form of $f(n)$ that produces the hump-shaped curve of the red or blue lines in Figure 1 can be a model of an Allee effect. This means that there are an almost unlimited number of potential Allee effect models, and several have been used in the literature. Only one model is described here to provide an example. In most cases, it is easiest to start with a model that has only negative density dependence and modify it to incorporate positive density dependence. For instance, one of the simplest mathematical models of population growth in continuously reproducing population is the logistic growth model, where the per capita growth rate is

$$f(n) = r\left(1 - \frac{n}{K}\right).$$

In logistic growth, the population growth rate dn/dt is a quadratic function so that the population grows most slowly when the population is very small and when the population is high and near carrying capacity. The *per capita* growth rate $f(n)$ declines linearly with population density (Fig. 1), and this negative relationship between per capita growth rate and population density shows that there is no Allee effect in this model. The fixed points of this model are at $n = 0$ and $n = K$, but $n = 0$ is an unstable equilibrium, which means that the population from any positive starting density will never go extinct but will increase and always reach the carrying capacity K. In logistic growth, $n = K$ is the only stable equilibrium so that from any starting density above or below K, the population will shrink or increase toward K (Fig. 2A).

We incorporate a strong Allee effect into this model by adding a parameter, a, representing the Allee threshold, $0 < a < K$, and the *per capita* growth rate now becomes

$$f(n) = r\left(1 - \frac{n}{K}\right)\left(\frac{n - a}{K}\right).$$

There are now three fixed points, $n = 0$, $n = a$, and $n = K$. The population growth rate is a cubic function of n, and the population grows slowly near the fixed points. Between 0 and a, the population growth rate is negative; above a, the population growth rate is positive. The *per capita* growth rate is quadratic (Fig. 1). The left-hand

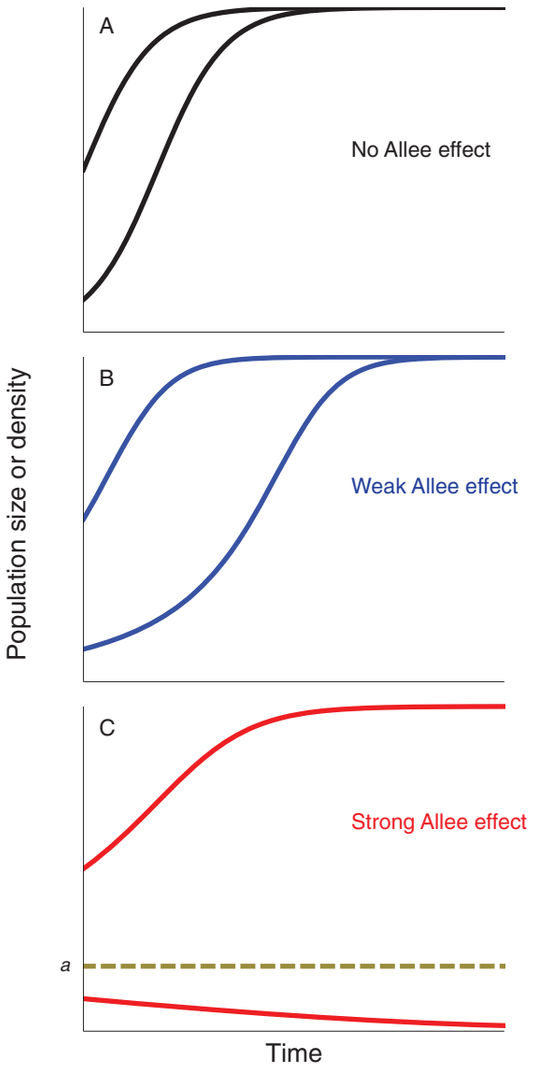


FIGURE 2 Trajectories of populations shown in Figure 1. (A) With no Allee effect, the population always grows toward carrying capacity, *K*. (B) With a weak Allee effect, the population also grows toward *K* but more slowly than with no Allee effect. (C) With a strong Allee effect, the population grows toward *K* if it is initially above the Allee threshold *a* but declines to extinction if it is initially above *a*.

side of this curve is the positive relationship between per capita growth rate and n that defines the Allee effect. Mathematically, the addition of the strong Allee effect stabilizes the $n = 0$ point (both $n = 0$ and $n = K$ are stable equilibria, and $n = a$ is unstable). Biologically, this means that population extinction is now possible. The population will decline to extinction if the starting density is below the threshold, a, and will grow to carrying capacity, K, if the population density starts above the threshold (Fig. 2C).

The same model can be used to incorporate a weak Allee effect by simply making $a < 0$. When a is negative, the population growth rate is still a cubic function of population density but never becomes negative (for positive, biologically reasonable values of n), so the population does not decline at low densities but always grows to carrying capacity (Fig. 2B). There is still a positive relationship between per capita growth rate and density, and dynamically this means that the population grows more slowly at low densities but will eventually reach carrying capacity as it does in the logistic growth model without an Allee effect.

Models of Component Allee Effects

MODELS OF MATE FINDING

One of the most commonly modeled component Allee effect mechanisms is mate finding. Usually to incorporate a mate-finding Allee effect, we need a sex-classified population model in which we have two variables, the number of males (M) and the number of females (F). The birth rate then depends upon the female mating rate $P(M, F)$, which is affected both by the number of females (F) in the population as well as the number of males (M). $P(M, F)$ can be any function that is zero if there are no males ($P(0, F) = 0$) and increases as the number of males increases, approaching 1 as the number of males gets very large relative to the number of females.

MODELS OF PREDATION

We can use a single-species framework to model predation if we can assume that the predator is not affected by the prey numbers, as would be the case for a generalist predator that has multiple other prey species. In this case, a component Allee effect is created in any system in which mortality is affected by density such that per capita mortality is higher when density is lower (in other words, a component of fitness, survival, is positively affected by density). A 2004 paper by Joanna Gascoigne and Romuald Lipcius shows that predation is a general mechanism that creates an Allee effect in prey species (Fig. 3). When predators consume prey, the consumption rate varies as the density of the prey changes. This is called a functional response and there are three types of functional response (Figs. 3A–C). A type I function response, used often for suspension feeders or predators that catch prey in traps, is a linear response in which consumption rate increases linearly with density. A type II functional response is a saturating function so that the consumption rate increases with prey density up to some density but then levels off. A type III

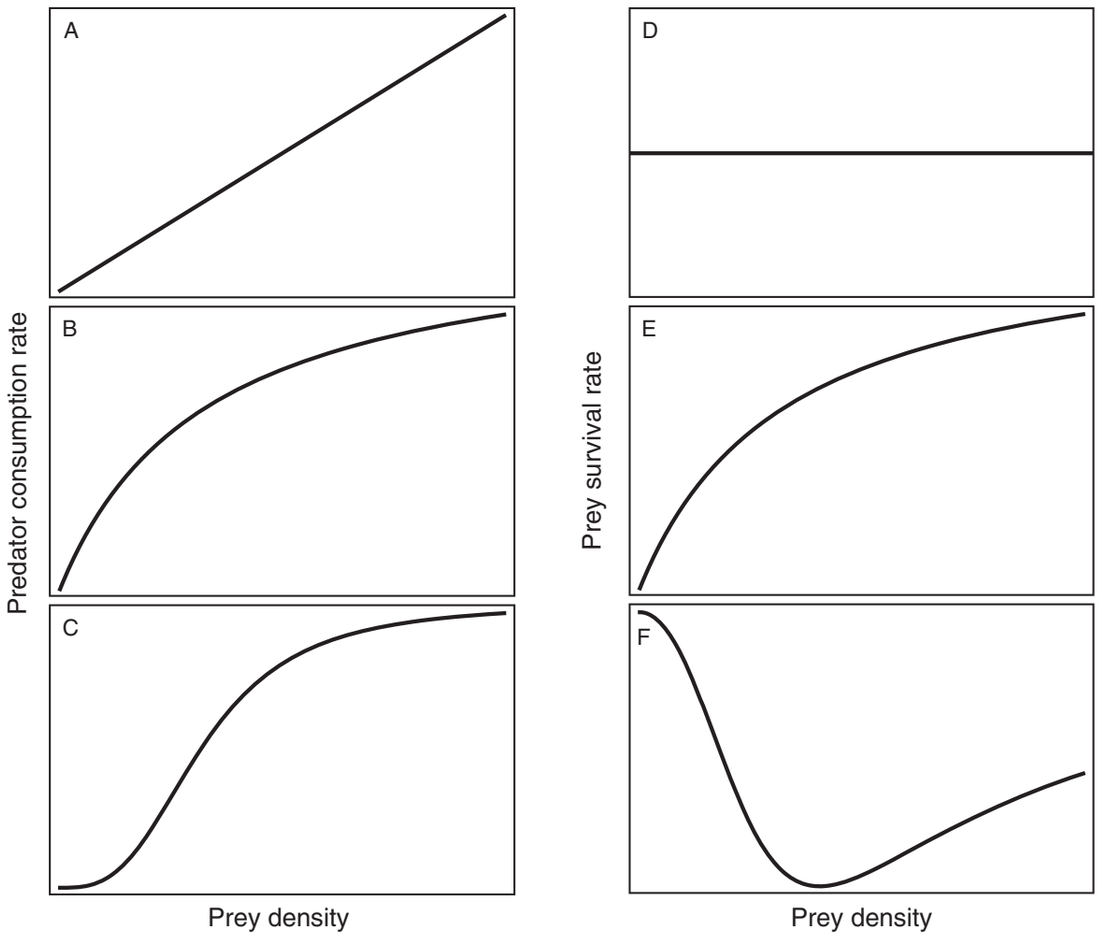


FIGURE 3 (A–C) The prey consumption rate of a predator following a type I, type II, and type III functional response, respectively. (D–F) The resulting prey survival rate. An Allee effect is created by a type II functional response but not by a type I or type III functional response. Adapted from Gascoigne and Lipcius (2004).

functional response has a sigmoid shape and is seen in predators that switch to different prey species when one species reaches low densities or in cases where small numbers of prey have refuges and can hide from predators at low densities. The resulting per capita survival of the prey as a function of density for the three types of functional response is shown in Figure 3D–F. The survival of the prey of a type I predator is does not vary with density (Fig. 3D). A type II response creates a positive relationship between prey survival and prey density (Fig. 3E), and a type III functional response produces a negative relationship between prey survival and density at low density (Fig. 3F). Hence, a type II functional response creates an Allee effect in the prey species, whereas a type I and a type III do not. The positive relationship between survival and density seen in the survival of prey with type III predators (Fig. 3F) is not considered an Allee effect because it occurs at higher densities. Predator density or abundance sometimes changes in response to prey density. This is called an aggregative response and can follow the same three types described above for functional response. An Allee effect caused by a type II functional response can be eliminated by including a type III aggregative response.

MODELING OTHER COMPONENT ALLEE EFFECTS

Predation and mate finding are the most commonly studied mechanisms for Allee effects, but there are many others and there have been many types of population models created to investigate them, including population genetics models, simulation models, competition models, predator–prey models, and metapopulation models. An extended discussion of all models of component and demographic Allee effects can be found in a 2008 book

about Allee effects by Frank Courchamp, Ludek Berec, and Joanna Gascoigne.

Metapopulation-Level Allee Effects

A classic metapopulation simply models the dynamics of patch occupancy. Each patch is occupied or not, and patches can become occupied at a colonization rate and become unoccupied at an extinction rate. In the classic formulation (the Levins model), colonization rate is proportional to the product of the fraction of occupied patches (p) and the fraction of unoccupied patches ($1 - p$). The per-patch colonization rate decreases linearly with p, whereas the per-patch extinction rate is constant and the metapopulation grows to a stable equilibrium fraction of patches occupied. It is possible, however, for situations to occur in which the per-patch colonization rate also drops with the number of occupied patches, creating what is termed an Allee-like effect, essentially an Allee effect that operates at the level of the metapopulation rather than at the level of local population. An Allee-like effect is usually analogous to a strong demographic Allee effect, creating a critical threshold fraction of patches occupied that needs to be exceeded for the metapopulation to persist. This occurs, for example, in parasite dynamics when infection of new hosts depends on there being more than a critical number of other infected hosts. Another example is in animals that form packs. When the number of packs drops below a critical threshold, not enough dispersers are produced and the rate of colonization or formation of new packs is reduced. Allee-like dynamics at the metapopulation level do not necessarily depend on Allee dynamics at the local population level, but Allee dynamics at the local level can lead to Allee-like dynamics at the metapopulation level.

DYNAMICAL IMPLICATIONS

Including Allee effects in population or metapopulation models has large consequences on how the population behaves. The most obvious consequence of a strong Allee effect is the introduction of a critical threshold for population level below which the population will go extinct. In stochastic models there is always some probability of extinction, but in stochastic models that include a strong Allee effect the probability of extinction declines sharply at the Allee threshold. In spatial models, this threshold is also spatial; the population must occupy an area larger than a critical threshold to remain viable (or a minimum proportion of patches in a discrete landscape). Also in spatial populations the rate at which a population spreads is reduced by both strong and weak demographic Allee effects. If habitat is patchy (discrete space), a strong Allee effect can produce pulsed invasions as local populations overcome Allee effects and colonize patches. In some cases, when local populations fail to overcome Allee thresholds this can prevent the population from expanding into available habitat patches (range pinning). In a continuous habitat, expansion of a population with a strong Allee effect can result in patchy distribution.

EVOLUTIONARY IMPLICATIONS

An Allee effect imparts a strong selection pressure and is responsible for many evolutionary adaptations. These include adaptations that cause the species to remain above the Allee threshold, like gregariousness, and adaptations that lower or remove the Allee threshold, for instance, adaptations that improve mate finding, like songs, pheromones, dispersal, mating synchronicity, and many others. Additionally, adaptations that decrease the frequency at which mating has to occur can be driven by Allee effects. These include traits as diverse as sperm storage and lifetime pair bonding, as well as adaptations that improve fertilization efficiency. We particularly expect species that are naturally rare to have evolved so as to minimize Allee effects. Practically, this means that we are most likely to be able to actually detect Allee effects in species that have become rare due to anthropogenic activities or are forced into low densities in manipulative experiments or lab situations as in many of Allee's original experiments.

CONSERVATION IMPLICATIONS

With the wide range of mechanisms described, it is likely that Allee effects or positive density dependence are very common and potentially have large effects on population dynamics and on the viability of populations. An Allee threshold, if one exists, can increase the likelihood of extinction of a rare species, and conservation biologists are likely to be most concerned about Allee effects in rare or endangered species or populations, especially those that are made rare by humans. Conservation of such species requires keeping the population above the Allee threshold. Invasive species are also affected by Allee effects. Sometimes the Allee threshold can be exploited for management if an invasive species needs only to be reduced to below its Allee threshold to be eradicated. For harvested species such as fisheries, a constant fishing effort does not create an Allee effect but does strengthen an existing Allee effect. However, using a constant yield

management plan generates a component Allee effect, since per capita mortality decreases as population size increases.

SEE ALSO THE FOLLOWING ARTICLES

Conservation Biology / Invasion Biology / Metapopulations / Population Ecology / Predator–Prey Models / Single-Species Population Models

FURTHER READING

Allee, W. C. 1931. *Animal aggregations: a study in general sociology.* Chicago: University of Chicago Press.

Allee, W. C. 1941. *The social life of animals,* 3rd ed. London: William Heineman.

Courchamp, F., L. Berec, and J. Gascoigne. 2008. *Allee effects in ecology and conservation.* Oxford: Oxford University Press.

Gascoigne, J. C., and R. N. Lipcius. 2004. Allee effects driven by predation. *Journal of Animal Ecology* 41: 801–810.

Stephens, P. A., W. J. Sutherland, and R. Freckleton. 1999. What is the Allee effect? *Oikos* 87: 343–61.

Stephens, P. A., and W. J. Sutherland. 1999. Consequences of the Allee effect for behaviour, ecology and conservation. *Trends in Ecology & Evolution* 14: 401–405.

Taylor, C. M., and A. Hastings. 2005. Allee effects in biological invasions. *Ecology Letters* 8: 895–908.

ALLOMETRY AND GROWTH

ANDREW J. KERKHOFF

Kenyon College, Gambier, Ohio

Allometry is the study of how organism body size affects various aspects of form and function. The allometric scaling of metabolic rate and other physiological processes explicitly links an individual organism's energy budget to changes in size. Several models of ontogenetic growth assume that organism growth trajectories arise from differential allometric changes in different components of metabolism.

A BRIEF OVERVIEW AND HISTORY OF ALLOMETRY

Individual organisms span an amazing size range. The ratio of the mass of a blue whale to that of a bacterium is approximately 10^{21}; that is, a blue whale is ten-thousand-million-million-million times heavier than the smallest bacteria. To put this ratio into perspective, it is similar to the mass ratio of the Moon to a typical human, that of a human to a single molecule of cytochrome-c oxidase (a protein that facilitates cellular respiration), and that of the known universe to our Sun! The scales of ecological interactions between species are yet broader, spanning over 30 orders of magnitude (powers of 10) in mass, from the smallest interacting microbes to the entire biosphere (which is estimated to weigh in at $\sim 1.8 \times 10^{19}$ g). Even among the more familiar land mammals, an elephant is almost six orders of magnitude (i.e., 10^6, or one million times) heavier than the smallest mouse. Thus, understanding biodiversity and the ecological complexity of life on Earth is, at least in part, a matter of understanding how life processes change across this range of scale.

While this observation seems obvious, it has important ramifications for almost every aspect of biology and ecology, because the size of an organism fundamentally affects its morphology and physiology, as well as its interactions with the physical environment and other species. For example, because they are "small," mice can easily climb a vertical surface and survive a fall from great height without injury. The same is clearly not true for humans, and a surprisingly modest tumble can fatally disable an elephant. The huge differences we see between small and large organisms are matched by equally intriguing constancies. For example, a mouse's heart beats 500 times per minute, while an elephant's only beats 28 times, but over their lifetimes, both will experience (on average) approximately the same number of heartbeats—about 1.5 billion, as will humans. Why 1.5 billion? Understanding how life must change with size is one of the keys to explaining these sorts of mysteries and to developing a more complete understanding of how nature works.

The term allometry (derived from the Greek *allos* = other, *metros* = measure) was coined by the biologists Julian Huxley and George Tessier in 1936. However, the changes in form and function that accompany changes in size have long intrigued scientists, dating back at least to Galileo's examination of the changing dimensions of bones and da Vinci's quantification of area-preserving branching in trees. One of the breakthroughs made by Huxley and Tessier was the standardized quantification of allometric relationships in the form of a power law,

$$Y = y_0 X^z,$$

where Y is a measure of some aspect of form or function, X is a measure of organism size (generally either length or mass), and y_0 (the allometric coefficient) and z (the scaling exponent) are constants that describe the proportionality of the relationship between size and the response. Box 1 presents more information on the mathematical aspects of allometric power laws.

BOX 1. EXTRACTING EIGENVALUES AND EIGENVECTORS FROM TRANSITION MATRICES

The allometric power law,

$$Y = y_0 X^z,$$

provides a simple mathematical model for the relationship between two measurable biological variables. If we take the logarithm of both sides of the allometric equation, we find an equation for a straight line relating log$Y$ to log$X$:

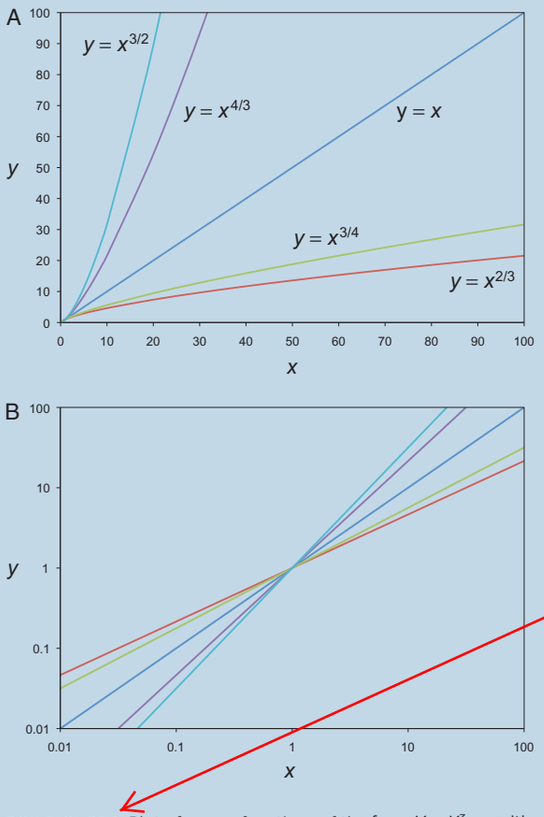
$$\log Y = \log y_0 + z \log X.$$

Thus, a power function relationship between two variables X and Y means that their logarithms are linearly related, with a slope equal to the exponent of the power function. Recall that a slope is defined as the "rise over run" or, in this case (since we are working with base-10 logarithms), the change in the magnitude of Y for every 10-fold change in X.

Looking at some graphs of power functions on both standard arithmetic and logarithmic scales can help to understand them better. The upper graph shows power functions with five different exponents. In all cases, the coefficient (y_0) is set to 1, for simplicity. Note that when the exponent is less than 1, the line climbs at an ever-decreasing rate, while when the exponent is greater than 1, it climbs at an ever-increasing rate. In all cases, the curves are not straight lines, that is, they are nonlinear, except when the exponent is equal to 1, for which $Y = X$. Finally, note that when viewed on an arithmetic scale, the five lines all appear to converge toward zero.

Now examine the same five power functions on logarithmic scales (lower graph), on which each equal increment is a power of ten. As we would predict from the derivation above, each function is linear, and since the coefficient a is always one, the lines only differ in their slopes, and they all cross at the point [1, 1]. The constant slopes indicate that for every 10-fold increment in X, Y changes 10^z-fold. Viewed on logarithmic scales, we can also see that the curves diverge on the small end as well as on the large end. These differences are not apparent on the upper graph, because they are compressed into the space between 0 and 1.

In a way, these two graphs represent different ways of looking at the world, and some of the differences are subtle. For example, the way in which the upper graph "minimizes" the differences between the power functions near the origin shows that arithmetic scales are sensitive to the units of measure: 1 mg is very different from 1 kg, but the space between 0 and 1 is always the same on the graph. Logarithmic scales, on the other hand, present proportional changes that are insensitive to the units of measure, because ten times larger is ten times larger, whether in mg or kg. This difference does not make one quantitative view superior to another, but it does mean that one or the other may be more appropriate, depending on the situation. Because allometric analyses are generally concerned with proportional changes, logarithmic scales are generally the appropriate choice.



BOX FIGURE 1 Plot of power functions of the form $Y = X^z$ on arithmetic (upper) and logarithmic (lower) axes.

Please check the figure caption here.

The importance of organism size for biological form and function and the long history of its study has led to applications of allometry in many different areas of biology. The portability of the mathematics of allometric power laws (Box 1) means that very different biological applications can take advantage of the same mathematical tools, but it is very important to recognize differences in biological context, especially when trying to develop generalized explanations for the relationships between size and biological form and function.

At least two fundamentally different approaches to allometry can be distinguished, based principally on the units of data used to develop the allometric relationship.

Comparative (also called evolutionary or interspecific) allometry studies how form or function changes across species that vary in size. Species are the unit of observation, and data generally consist of species mean values of adult size and the biological variable of interest. The physiological allometry relating basal metabolic rate to body mass in mammals (see "The Allometry of Metabolism," below) is a classic example. In contrast, intraspecific (or ontogenetic) allometry examines variation across individuals of a single species that vary in size, either developmentally or ecologically. The patterns that arise in different sorts of allometric studies do not always agree, in part because the observed differences in both size and biological form and function reflect different combinations of evolutionary, physiological, ecological, and developmental processes.

THE ALLOMETRY OF METABOLISM

Early studies of allometry were concerned primarily with patterns of relative growth, such as the relationship between brain size and body size among mammals. However, from the beginning, biologists applied similar ideas to patterns in the physiology, ecology, and behavior of organisms. During the 1930s, Max Kleiber and other animal physiologists brought together careful measurements of basal metabolic rates (B) for a wide variety of mammals. The data displayed a very tight relationship with adult animal mass (M), with a scaling exponent of approximately 3/4, *i.e.,* $B = b_0 M^{3/4}$. In the decades since, subsequent studies of very different taxa, from fish to birds to insects to plants, have confirmed similar allometric patterns for almost all multicellular (and perhaps even unicellular) organisms (Fig. 1). While there has been decades of considerable debate concerning the "universality" of the value of the metabolic scaling exponent and a number of controversial attempts to explain its origin, for our purposes it is only important to note two facts of metabolic allometry, neither of which depend on the exact value of the exponent or its basis.

First, most (but perhaps not all) analyses, whether interspecific, intraspecific, or ontogenetic, find relationships with exponents less than 1. This implies that as organisms get larger, their mass specific rate of metabolism decreases, since, for example, if $B = b_0 M^{3/4}$, $\frac{B}{M} = b_0 M^{3/4} M^{-1} = b_0 M^{-1/4}$. Thus, even though organisms large and small share a common biochemistry and cellular structure, the cells of larger organisms run their biochemistry at slower rates than the cells of smaller organisms.

Second, in addition to direct measurements of respiration, a variety of other biological rates, including rates

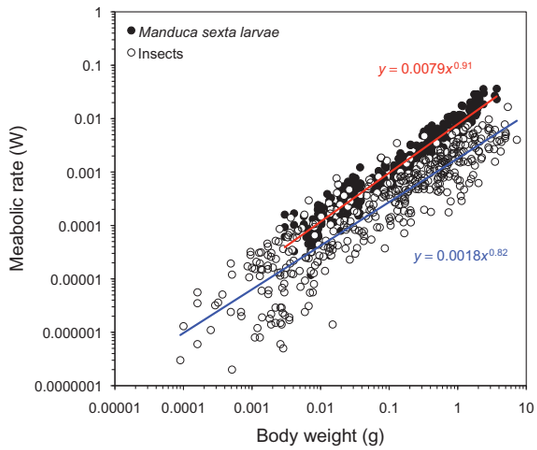


FIGURE 1 Allometric scaling of metabolic rate as a function of body mass for insects showing evolutionary (interspecific) allometry across 392 species of adult insects (open circles with blue line, compiled from the literature by Chown et al., 2007, *Functional Ecology* 21: 282–290) as well as ontogenetic (intraspecific) allometry measured daily during development for N individual larvae of the tobacco hornworm, *Manduca sexta* (solid circles and red line, unpublished data from A. Boylan, H. Itagaki, and A. Kerkhoff). Note that the scaling exponents differ between the two relationships. Differences in the height of the relationship (the scaling coefficient) likely derive from differences in temperature. The adult insect data were all corrected to 27 °C, while the *Manduca* larvae were reared at 27 °C.

of ingestion, excretion, reproduction, somatic growth, and even mortality have been shown to share similar patterns of allometric scaling. This implies a principle of similitude (or similarity) linking these different physiological and demographic processes. That is, if different processes scale similarly with body size (i.e., with the same exponent), their ratio is approximately constant, or at least does not vary systematically with the size of the organism. It is exactly this kind of similitude that produced the example above of 1.5 billion heartbeats per lifetime in mammals. Heart rate, like mass-specific metabolic rate, scales with an exponent of about $-1/4$. Average lifespans, which are the inverse of mortality rates, scale with an exponent of 1/4. The number of beats per lifetime is simply the product of these two relationships, and since $X^{-1/4} X^{1/4} = X^0$, we can predict, on average, that all mammals experience about the same number of heartbeats over the course of their life, regardless of their size.

The generality of metabolic scaling and the similitude that it shares with so many other biological rates and times has broad implications for the ecology and life history of organisms, the dynamics of populations, and the functioning of ecosystems. Some of these implications are

explored in greater depth elsewhere in this volume. Here, we explore their implications for modeling the growth of individual organisms.

ALLOMETRIC MODELS OF ANIMAL GROWTH AND PRODUCTIVITY

Most models of organismal growth begin with the balanced growth assumption, which applies the first law of thermodynamics (conservation of mass and energy) to biological systems. That is, any change in the size of the animal must result from the balance of material and energetic inputs and outputs, and

$$\text{Growth} = \text{Ingestion} - \text{Egestion} - \text{Excretion} - \text{Respiration} - \text{Reproduction}.$$

This basic balanced growth model is cast entirely in terms of rates. Since most biological rates exhibit allometric scaling, growth trajectories represent the balance of allometric changes of the various input and output rates.

Typically, growth trajectories have a roughly sigmoid pattern, with accelerating growth early in development (i.e., near their initial mass, m_0) followed by a progressive slowing of growth as the organism approaches its asymptotic mass, M (see Fig. 2). In some organisms (e.g., insect larvae), the curve appears nearly exponential and then abruptly ceases without any gradual slowing. The sigmoidal pattern is similar to the pattern of logistic growth observed in natural populations growing under some sort of resource limitation. A generalized mathematical function that captures the pattern of sigmoid growth is

$$\frac{dm}{dt} = am^{\alpha} - bm^{\beta},$$

where m is organism mass (and dm/dt is thus the growth rate). To conform to a sigmoid pattern of growth, the two allometric relationships (with coefficients a and b, and exponents α and β, respectively) must have the additional constraints that $a > b$ (which assures positive initial growth at $m = m_0$) and $\alpha < \beta$ (which leads to a cessation of growth at $m = M$). The asymptotic adult mass of the animal, M, is the point at which $0 = \frac{dm}{dt} = am^{\alpha} - bm^{\beta}$; thus, $M = \left(\frac{a}{b}\right)^{\frac{1}{\beta-\alpha}}$. Likewise, the inflection point, i.e., the size at which the growth rate is maximized, is a constant fraction of the asymptotic mass, $\left(\frac{\alpha}{\beta}\right)^{\frac{1}{\beta-\alpha}}M$. Note this may, however, be less than the initial size m_0, in which case maximum growth rate occurs at the initial size.

Depending on the parameter values assigned (or derived), this general equation (which is sometimes called the Pütter equation) corresponds to several prominent models of ontogenetic growth. For example, $\alpha = 1$ and $\beta = 2$ gives the classical logistic growth model.

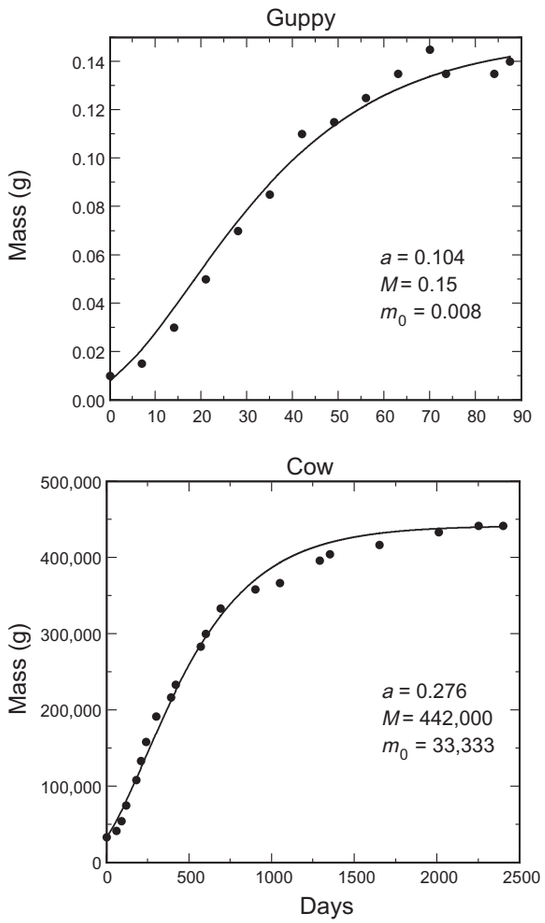


FIGURE 2 Ontogenetic growth trajectories and fitted model parameters for two vertebrates, the guppy and the cow (redrawn from West et al., 2001, *Science* 413: 628–631). The model was the form of the Pütter equation assumed by West et al., i.e., with $\alpha = 3/4$ and $\beta = 1$. Despite the vast evolutionary divergence and differences in adult size (M) and size at birth (m_0), both taxa show a common sigmoidal pattern of growth. According to the model, differences in the parameter a principally reflect differences in the scaling coefficient of the metabolic allometry (b_0) between fish and mammals.

While the logistic model sometimes provides a reasonable fit to growth trajectory data, it is difficult to assign biological meaning to the model parameters based on the balanced growth assumption, because most of the biological rates and times involved tend to exhibit scaling exponents less than 1. This entry will review three approaches that differ not just in parameter values but also in the method by which those parameter values are derived and thus assigned biological meaning. For an example of another approach based on many of the same principles, see the discussion of dynamic energy budgets elsewhere in this volume.

In 1947, Ludwig von Bertalanffy published a model of ontogenetic growth based on the balance of anabolic processes (represented by am^{α}) and catabolic processes (bm^{β}). Further, he assumed that anabolic processes were limited by the surface area over which organisms assimilate resources, and that the costs of catabolism were directly proportional to the mass of the animal, i.e., $\beta = 1$. Because mass is proportional to volume (length cubed), while areas are length squared, the surface area assumption leads to $\alpha = 2/3$. Bertalanffy based this assumption on an interpretation of the metabolic allometry that pre-dates Kleiber's work and still persists today. Surface areas are thought to be a critical constraint on the allometry of metabolism because they limit the rate at which organisms can assimilate resources (e.g., across the gut surface) or eliminate wastes (e.g., heat dissipation in homeotherms). The resulting model $dm/dt = am^{2/3} - bm$ has been applied widely, especially to the growth of marine fishes. Subsequently, Bertalanffy recognized different growth types corresponding distinguished by whether their metabolic rates scaled with surface area ($\alpha = 2/3$), mass ($\alpha = 1$), or in some intermediate way (e.g., $\alpha = 3/4$). In all cases, he assumed that the exponent α reflected the scaling of metabolic rate.

In 1989, Michael Reiss developed an alternative interpretation, still based on the structure of the Pütter equation. He assumed that, instead of the balance of anabolism and catabolism, the first term represents the scaling of resource input (which is Ingestion − Egestion in the balanced growth model), while the second term represents the metabolic cost of living, which couples respiration and excretion. While Reiss recommends parameterizing the resulting model based on empirical allometric relationships for particular taxa, he utilizes $\alpha = 2/3$ and $\beta = 3/4$ as a general solution, based on a broad survey of empirical allometries for ingestion rates and metabolic rates, respectively. Note that although the forms of their models are quite similar, Bertalanffy assumes that metabolic allometry is reflected in the first term, while Reiss assumes it is reflected in the second term. Thus, even if the two models provide reliable fits to empirical growth data, their biological interpretation is very different. More generally, distinguishing between the two models cannot be accomplished simply by comparing their fit-to-growth trajectories alone. Instead, their underlying assumptions must be addressed.

In 2001, West, Brown, and Enquist presented a model that while conforming to the same overall form, had yet another derivation. They began with the assumption that the total metabolic rate of an organism (B) is simply the sum of the energy devoted to growth (i.e., the synthesis of new biomass) plus the energy devoted to maintaining existing biomass,

$$B = E_m \frac{dm}{dt} + B_m m$$

where E_m is the energy required to synthesize a unit of biomass (e.g., in $\mathrm{J\,g^{-1}}$) and B_m is the metabolic rate required to maintain a unit of biomass (e.g., in $\mathrm{W\,g^{-1}}$). Rearranging to solve for the growth rate, and taking into account the allometry of metabolic rate (i.e., $B = b_0 m^{\alpha}$), they arrive at a Pütter-style model,

$$\frac{dm}{dt} = am^{\alpha} - bm,$$

where

$$a = \frac{b_0}{E_m} \text{ and } b = \frac{B_m}{E_m}.$$

Thus, a is the ratio of the metabolic scaling coefficient (b_0) to the cost of biomass synthesis, and b provides a measure of tissue turnover rate. In the original derivation, they made the additional assumption that $\alpha = 3/4$, as in the interspecific metabolic allometry documented by Kleiber, but later versions were generalized to any metabolic scaling exponent. While their model is thus almost identical in form to Bertalanffy's, the important advance is that they provide an unambiguous (and thus testable) biological interpretation not just of the exponent (α) of the growth model (i.e., it should match the observed exponent for metabolism) but also of the two coefficients (a and b), which are related to the mass-specific metabolic parameters (b_0 and B_m) and the energetic costs of biomass synthesis (E_m). Although estimates of these parameters from the original analysis of their growth model appear to be biologically reasonable, further formal tests of the model are required.

These three examples serve to illustrate three points concerning ontogenetic growth models and their relation to allometry. First, while all of the models are based on the balanced growth assumption, they make very different assumptions about what the different terms in the model represent, leading to different applications of allometric principles. For example, whereas Reiss expects metabolic allometry to be reflected in the second term, the other two models associate it with the first term. Second, because models based on very different assumptions can take on very similar or even identical mathematical forms, distinguishing between the models as alternative hypotheses cannot be accomplished by comparing their ability to fit growth trajectory data alone. Instead, comparative studies must address the underlying assumptions of the models and their

ancillary predictions to assess whether the models capture growth trajectories for the right reasons. Finally, despite a rather long history of employing allometric principles to the study of ontogenetic growth, open questions remain that require careful mathematical and biological reasoning and the confrontation of abstract models with carefully collected data. Further developments in ontogenetic growth modeling have included the effects of temperature and the elemental stoichiometry of organismal growth, and ongoing experimental approaches have begun to address model assumptions by taking measurements of growth, assimilation, and metabolism on the same organisms.

ALLOMETRIC INSIGHTS ON ANIMAL GROWTH AND LIFE HISTORY

The discussion above ignored one component of the balanced growth model that is critical both for the persistence of organisms and for understanding ontogenetic growth models in an evolutionary context: reproduction. Including this fundamental biological process in the model requires mathematical modifications, the details of which depend on whether the organism in question exhibits determinate or indeterminate growth. For determinate growth (as observed in most mammals and birds), the size at maturity (first reproduction) corresponds to the asymptotic size, M. All growth takes place before reproduction begins and reproductive investments are either assumed to be supported by metabolic scope (represented by temporal variation in a) or to tradeoff against the costs of maintenance, e.g., tissue turnover, b. For indeterminate growers, like many fish, perennial plants, and some invertebrates, the issue is more complicated. From birth to the age (and size) at maturity, the models outlined above apply directly. However, once the organism begins reproducing, the parameters of the equations need to be adjusted to reflect the allocation to reproduction vs. growth, because any materials and energy allocated to reproduction are by definition unavailable for growth. Thus, indeterminate growers may slow or even cease growth well below their asymptotic size.

Although the dynamics of life history can complicate the modeling of growth, many aspects of life history display allometric scaling, themselves. Based on principles of similitude like those described above, Eric Charnov and others have focused on particular dimensionless numbers in life history that are independent of size. For example, the ratio of turnover rate (b in the West et al. and Bertalanffy growth models, above) to adult mortality rate (Z) have both been shown to scale across species within a taxonomic group such that their ratio (b/Z) is invariant with respect to size. This invariance is essentially the same as the invariant number of heartbeats described above; both represent an invariance in the average total lifetime energy flux observed of an organism. Although this invariance is itself remarkable, it is perhaps even more interesting that different taxonomic groups exhibit different average values. For example, the estimate of b/Z for fish (0.13) is an order of magnitude lower than that of mammals (14), and it is likely even higher for birds, which at a given size have both a faster metabolic rate (higher b) and a longer lifespan (lower Z) than mammals.

Life history invariants like b/Z provide a means of extending models of ontogenetic growth to higher-order, population-level processes. By combining a model of ontogenetic growth with life history invariants and assuming a stable age distribution, it is possible to predict how the production to biomass ratio (P:B) of whole populations changes with average adult body size (Fig. 3). Moreover, measured differences in the values of the life history invariants correspond to observed quantitative differences in the scaling of P:B between taxonomic groups, e.g., mammals and fish.

The quantitative regularity of allometric scaling relationships makes them a powerful predictive tool for

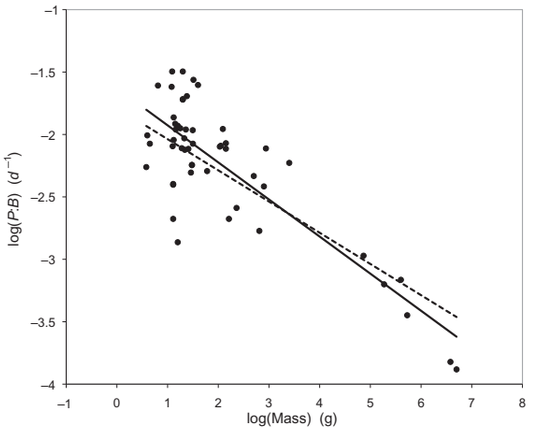


FIGURE 3 Population-level scaling of the ratio of production to biomass (P:B) as a function of average adult size for mammals (redrawn from Economo et al., 2005, *Ecology Letters* 8: 353–360). The dashed line is the regression fit to the data, and the solid line is the theoretical prediction obtained from combining an ontogenetic growth model with principles drawn from the study of life history invariants, principally the invariance of maintenance costs of tissue turnover (b) relative to mortality rate (Z).

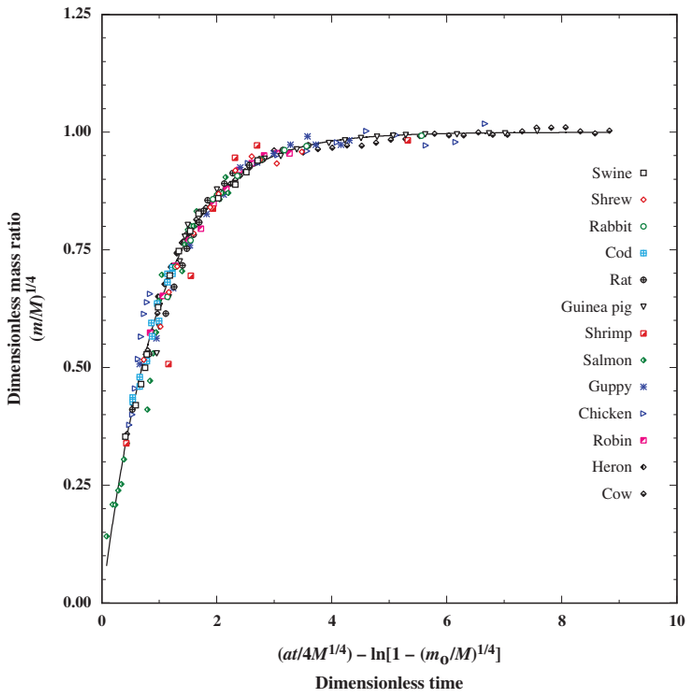


FIGURE 4 Plot of the "universal growth curve" derived by West et al. The dimensionless mass and time variables calculated for a wide variety of taxa exhibiting both determinate and indeterminate growth all fall quite near the same curve, $r = 1 - e^{-\tau}$, which is shown in the solid line. According to the West et al. model, the dimensionless mass ratio $r = \left(\frac{m}{M}\right)^{\frac{1}{4}}$ is the proportion of metabolic power used for maintenance and other activities not contributing directly to growth.

organizing and understanding variation in biological form and function. Part of the power of this point of view is that any particular animal (or species of animal) can be seen as a scale model of any other; Kleiber's famous relationship $B = b_0 M^{3/4}$ tells us literally how to use a mouse as a scale model for an elephant, energetically speaking. Likewise, the application of allometric principles to patterns of ontogenetic growth implies that individual growth trajectories are simply examples of a more general process that has been allometrically rescaled by the particular evolved physiology and ecology of the taxa being examined. In their 2001 paper, West, Brown, and Enquist make this rescaling explicit when they renormalize the parameters of the growth trajectories for a wide variety of species to show that animals as different as cod, shrimp, chicken, and dog all fall along a single, rescaled growth curve (Fig. 4). The scaling properties of allometries and ontogenetic growth models may represent deep biological symmetries—pervasive "rules" that generate, guide, or constrain biological form and function. In addition to providing powerful predictive tools, they suggest an elegant unity underlying the fascinating diversity of life.

SEE ALSO THE FOLLOWING ARTICLES

Energy Budgets / Metabolic Theory of Ecology / Stoichiometry, Ecological

FURTHER READING

Bonner, J. T. 2006. *Why size matters*. Princeton: Princeton University Press.
Calder, W. A. 1984. *Size, function, and life history.* Cambridge, MA: Harvard University Press.
Charnov, E. L. 1993. *Life history invariants: some explorations of symmetry in evolutionary ecology.* Oxford: Oxford University Press.
Kleiber, M. 1961. *The fire of life*. New York: Wiley.
McMahon, T. A., and J. T. Bonner. 1983. *On size and life.* New York: Scientific American Books, W. H. Freeman.
Niklas, K. J. 1994. *Plant allometry: the scaling of form and process.* Chicago: University of Chicago Press.
Peters, R. H. 1986. *The ecological implications of body size.* Cambridge, UK: Cambridge University Press.
Reiss, M. J. 1989. *The allometry of growth and reproduction.* Cambridge, UK: Cambridge University Press.
Schmidt-Nielsen, K. 1984. *Scaling: why animal size is so important.* Cambridge, UK: Cambridge University Press.
West et al., 2001. *Science* 413: 628–631.
Whitfield, J. 2006. *In the beat of a heart: life, energy, and the unity of nature.* Washington, DC: Joseph Henry Press.

ANIMAL DISPERSAL

SEE DISPERSAL, ANIMAL

APPARENT COMPETITION

ROBERT D. HOLT
University of Florida, Gainesville

Apparent competition is an indirect negative interaction between species mediated through the action of a shared natural enemy. The concept of apparent competition illuminates how natural enemies at times constrain the species richness of communities but at other times help maintain diversity. It is an integral part of any community theory focused on predators and their prey, parasites and their hosts, herbivores and the plants they consume, or entire food webs. In applied ecology, apparent competition can influence conservation risks, the success of invasive species, epidemiological patterns, and the efficacy and dangers of pest control. Finally, apparent competition exemplifies the general theme that ecological communities are not a haphazard assemblage of species, each interacting separately with the external environment, but instead exhibit complex and at times surprising chains of interactions.

ECOLOGICAL THEORY CAN HELP UNIFY ECOLOGICAL UNDERSTANDING OF DISPARATE SYSTEMS

Natural historians glory in life's rich diversity. Ecological theory can help tease out commonalities among disparate systems and thus help unify understanding. What do the following real-world stories have in common, and how can theory help explain the patterns they reveal?

Case Studies: Ecological Puzzles from Studies of Interacting Species

DO RABBITS "EAT" VOLES?

In the Grampian Mountains of Scotland, the native water vole, *Arvicola amphibius,* resides in lush vegetation along streams and other water bodies. During the eighteenth century, the European rabbit, *Oryctolagus cuniculus,* colonized the area, and its warrens of deep burrows are mainly found in dry fields of short grass. The two species live in different habitats and have different diets, and the rabbit invasion had no obvious impact on the water vole. Sadly, the water vole has plummeted in abundance over the last 50 years from millions to a few hundred thousand across the entire UK. In the Grampians it has disappeared from sites near upland fields containing rabbits. If the two species do not compete, how can the presence of one affect the fate of the other?

ANOTHER PUZZLE IN AN INVASION

In the United Kingdom, the native red squirrel, *Sciurus vulgaris*, once widespread, has been largely supplanted by the introduced grey squirrel, *Sciurus carolinensis*. The two squirrel species overlap in diet and habitat and so probably compete. But there are puzzling aspects of the decline: sometimes the red squirrel disappears from a locale even before the grey squirrel has built up its numbers. Is this just a straightforward story of competitive exclusion?

COUNTERINTUITIVE EFFECTS OF PROTECTING WILDLIFE

In the dry savanna of northern Kenya, humans, livestock, and wildlife have coexisted for millennia. To boost ecotourism, there have been localized shifts from cattle ranching to wildlife conservation. Some species, in particular the plains zebra, *Equus quagga burchellii*, increased greatly post-protection, but other ungulates such as hartebeest (*Alcelaphus buselaphus*) have severely declined. Are these declines driven mainly by competitive interactions among these large mammalian herbivores, or something else?

AN ENDANGERED ISLAND SPECIES

Feral pigs (*Sus scrofa*) were introduced into the California Channel Islands, which later saw abrupt crashes toward near-extinction of an endemic predator, the island fox, *Urocyon littoralis*. How did the pig introduction endanger the fox—could it for instance be habitat degradation driven by the destructive rootings of the pigs?

Simple Models of Apparent Competition

NATURAL ENEMIES ARE SIGNIFICANT FACTORS IN MOST SPECIES' LIVES

At first glance, several of these case studies are consistent with the hypothesis that species are directly competing—and this may be true. But to understand what forces drive these systems, it turns out one must consider species or populations beyond those that at first glance seem to be the main players—and all these empirical patterns reflect apparent competition.

Communities are complex webs of interacting species, affecting each other not just one on one but via interlocking chains of indirect interactions mediated through other species. Most species suffer from natural enemies—a term broadly encompassing predators, herbivores, parasites, pathogens, and indeed any species that to make a living inflicts harm on other species, taking resources (energy and nutrients) from living organisms so as to survive or reproduce itself. Even the fiercest top predator—lion,

tiger, or bear—can be laid low by an infectious disease agent, and the smallest quasi-organisms of all—viruses—themselves can be attacked by other viruses or consumed when their infected hosts fall prey to natural enemies.

Many natural enemies are generalists, exploiting more than just one victim species (prey or host). The level of damage imposed on any particular prey or host species (as assessed by reduced growth rates, fitness, or abundance) may depend upon the availability, traits, and productivity of alternative prey or hosts. In other words, there is an indirect interaction between these species. Formally, in a mathematical model, an indirect interaction exists between species I and III, mediated through II, if the dynamical equation describing growth rates in I contains a variable referring to II (e.g., abundance), and the equation for II has a variable referring to III. Changes in III lead to changes in II, which in turn affects I. When an indirect interaction between two victims mediated by a natural enemy is negative, it is called apparent competition. This term was coined by the author in 1977 because ecological patterns that appear to be due to competition for resources, such as nonoverlapping spatial distributions, can also emerge from impacts of a shared natural enemy.

A COMPARISON OF EXPLOITATIVE AND APPARENT COMPETITION

Figure 1 depicts community modules corresponding to exploitative competition and apparent competition. Each node represents a species, and arrows represent directions of effects. On the left, two natural enemies are consumers of the same resource, and any resource gathered by species A is thus unavailable to species B, which therefore suffers a reduction in its growth rate. Apparent competition is a mirror image of this familiar indirect interaction. Victim species A has a positive effect on the abundance or activity of a natural enemy, and because the natural enemy has a negative effect upon species B, species A indirectly has a negative effect upon species B.

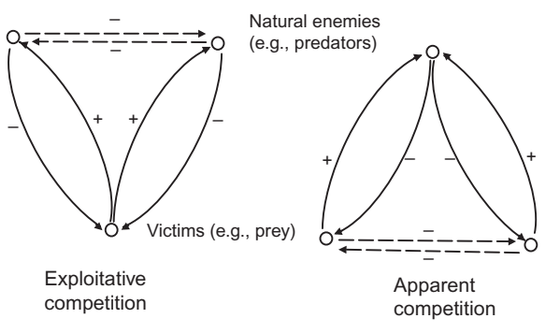


FIGURE 1 Exploitative and apparent competition. Each node is a species. Solid arrows are signed direct interactions; dashed arrows, emergent indirect interactions. Higher species in the figure are higher in the food web. On the left, two consumers (e.g., predators) share a resource. Multiplying the signs, each consumer has a negative indirect effect upon the other—competition for resources. On the right, two victim species (e.g., prey) share a natural enemy (predator). Again, multiplying the signs shows that each victim indirectly harms the other, via their shared natural enemy.

A GRAPHICAL MODEL OF APPARENT COMPETITION

A simple graphical model exploring this process provides one step towards a formal theory of apparent competition. Assume a predator of abundance P has an instantaneous population growth rate dependent only on availability of two prey species, of abundance R_1 and R_2:

$$\frac{1}{P}\frac{dP}{dt} = F_P(R_1, R_2). \qquad (1)$$

Assume the system reaches equilibrium, so $F_P(R_1, R_2) = 0$. We plot this (the predator zero-growth isocline) as a curve on a graph with axes of prey abundance (Fig. 2A). Outside the isocline, the predator grows; inside, it declines.

Assume each prey species when alone can sustain the predator. The predator reaches equilibrium when prey abundances match the open circles in the figure. Now assume an equilibrium exists with both prey species present. This equilibrium must lie on the predator isocline, for instance at the closed circle. If we compare the equilibrial abundance of each prey when alone to its abundance when the two coexist, note that at the joint equilibrium each prey species is depressed in abundance. The little arrows in the figure show that adding a small amount of either prey to the system permits predator growth. Mathematically, this is equivalent to saying the predator isocline has negative slope. Thus, each prey species indirectly depresses the equilibrial density of the other because each benefits the predator, increasing its numbers and thus predation upon the alternative prey.

A SIMPLE RULE FOR DOMINANCE IN APPARENT COMPETITION

But maybe no equilibrium exists with both prey species. Imagine the predator depresses prey species i to a level where it experiences only density-independent per capita growth at rate r_i and predation at rate a_iP (a_i is the attack rate, and P predator abundance). The net per capita growth of species i is $r_i - a_iP$. For equilibrium, the predator must settle to $P_i^* = r_i/a_i$ (the asterisk indicates equilibrium, and the index on P indicates just prey i is present). Suppose prey 1 is at equilibrium with the predator, and prey 2 tries to invade. Invasion succeeds only if $r_2 - a_2P_1^* > 0$, which implies $P_2^* > P_1^*$. But if this

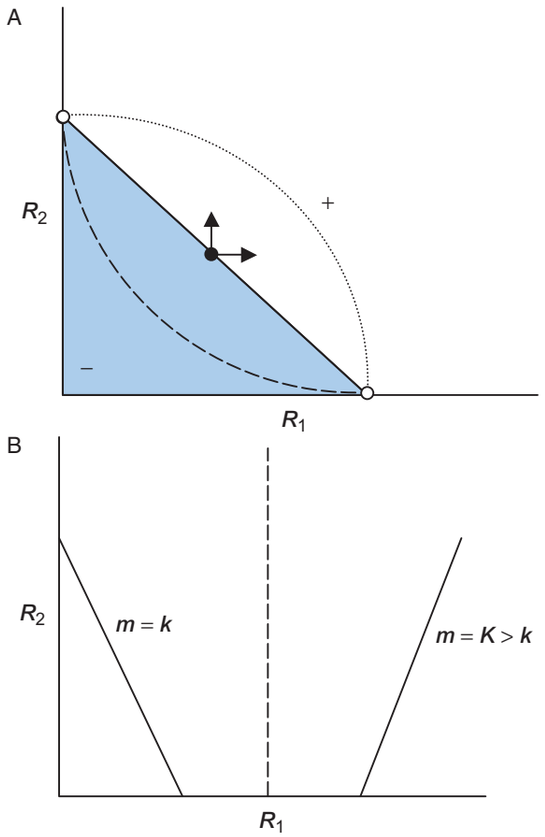


FIGURE 2 A graphical model of shared predation. (A) The solid straight line is the zero growth isocline of a predator. Between this line and the origin, the predator declines in abundance because of a shortage of food; outside the line, it thrives and grows in numbers. The small arrows show that increases in either prey leads to predator growth, and this leads to a negative isocline slope, and apparent competition (compare open circles to closed circles; see text). Apparent competition is even stronger if the predator thrives better on a mixed diet (dashed line), and weaker if it does not benefit from the mix (dotted line). (B) The model is for a food-limited predator with a saturating response to each of two prey:

$$\frac{1}{P}\frac{dP}{dt} = F_P(R_1, R_2) = \sum_{i=1}^{2} \frac{a_i b_i R_i}{1+a_1 b_1 R_1 + a_2 b_2 R_2} - m,$$

where P and R_i are respectively the abundances of the predator and prey species i; a_i, b_i, h_i are attack rates on prey i, the benefit the predator gains per prey caught, and handling times; and m, the predator death rate. We assume one prey is better than the other (as measured in b/h, benefit gained per unit handling time). If the predator is a generalist for all prey densities, at low abundances it benefits from each prey, so the isocline has a negative slope, but at high abundances, it can waste time on the poor prey (see text). At low predator mortality (the line indicated with $m = k$), the two prey experience apparent competition, but at higher mortality (marked by $m = k$), the isocline has a positive slope, so the indirect interaction between prey is (−, +).

holds, then when prey species 2 in turn is present at equilibrium, and prey species 1 tries to invade, it cannot—predation upon it exceeds its intrinsic $>k$ growth rate.

If attack rates are constant but intrinsic growth rates and predator densities vary over time, one can replace the r's and P's with time averages, and this conclusion still holds. In other words, if the predator is the only regulatory factor limiting each prey species, one expects to see exclusion of one prey species by the other, and that prey which sustains higher average predator abundance wins in apparent competition. Shared predation can constrain prey species diversity.

All the Above Ecological Puzzles Involve Apparent Competition.

Apparent competition arises in many different ecological systems. The solution to each ecological puzzle above turns out to involve apparent competition. Each illuminates different features and realistic complications of apparent competition and points out directions for theory development.

RABBITS DO NOT EAT VOLES—BUT MINK DO

In the first example above, there is another invasive species that also moved into Scotland, but well after the rabbits—the American mink, *Neovison vison*. This species is a smart, voracious predator, and it can follow the water vole into all its normal refuges. Minks do consume rabbits, but they cannot readily eliminate them from their protective warrens. In short, mink numbers get boosted by a species that they cannot control—the rabbit—which then permits the mink to impose very heavy mortality on a more vulnerable species—the water vole. Rabbits in no way directly compete with water voles, but nonetheless by sustaining a predator—the mink—they have a strong, albeit indirect, deleterious impact upon the persistence of water voles. This appears to be an excellent (if sad) example of apparent competition in action.

The water vole–rabbit–mink interaction matches expectations of our simple model. Mink populations respond strongly to increased food availability, so are food-limited. Rabbits have notoriously high fecundity and can reach high abundance; they have a high intrinsic growth rate. Many rabbits are protected from mink predation in their warrens and so have an average lower (but nonzero) attack rate than the water voles. Rabbits, we can reasonably infer, have a higher r/a than do water voles and so should tend to dominate in apparent competition. This appears to be what happens—the water vole is excluded from sites near rabbit warrens.

Spatial Processes and Apparent Competition This example can be used to glean other insights that have been explored in theoretical models. The theory of apparent competition has been extended from closed to open

communities, comprised of habitat patches linked by dispersal. If different prey species occupy different habitats (and so do not compete), their dynamics may be nonetheless coupled by predator movement. It turns out that alternative prey species can coexist, if prey species segregate among habitats and predator mobility is constrained. Water voles and rabbits do live in quite different habitats. The reason their fates are joined at all in the Scottish landscape is that minks are highly mobile. The farther a water vole population is located from rabbit populations, the less likely it is that mink (sustained by feeding on rabbits) will wander through. Remnant populations of water voles in the Grampians tend to be those at some distance from rabbit habitat, presumably because they suffer less "spillover" predation.

A critical dimension of apparent competition is thus the nature of spatial processes connecting different habitats. Natural enemies are often mobile, permitting apparent competition to act at a distance. This effect has been studied in laboratory microcosms consisting of patch arrays containing two moth species. Arrays were set up so that moths lived in different patches and did not directly compete, but a mobile wasp parasitoid could move freely among them. The wasp coexisted just fine with either host when alone, but the three-species combination collapsed, and one host species went extinct from elevated parasitism—apparent competition. In like manner, movement of prey across space can subsidize predators in a community, permitting them to more effectively limit resident prey.

WITH A LITTLE HELP FROM MY FRIENDS

Apparent competition is implicated in the conservation risks experienced by many endangered species, including the replacement of red by grey squirrels. The squirrels do seem to compete for resources, and the grey squirrel seems more effective at using acorn resources. Hence, grey squirrels can potentially have a higher carrying capacity and so may dominate in competition for shared resources. But on top of this, the grey squirrel harbors a poxvirus (SQPV)—an infectious disease agent to which they are seemingly immune but to which the red squirrel is highly vulnerable. Mathematical models suggest this shared pathogen greatly speeds up the demise of the red squirrel, even in advance of an increase in grey squirrel numbers.

Apparent Competition Mediated by Parasitism This example illustrates the potential importance of pathogens as conduits of apparent competition. This is a very important issue for conservation and public health. In California, the invasive pathogen *Phytophthora ramorum* is a highly generalized natural enemy, inflicting serious damage on many native tree species. The California bay laurel (*Umbellularia californica*) is a heavy producer of pathogen spores but is not itself greatly harmed. The spillover of spores from this species onto more vulnerable species such as tanoak (*Lithocarpus densiflorus*) leads to high mortality. Because trees compete, as vulnerable trees decline the bay laurel increases and hammers its competitors via the pathogen even harder.

Asymmetric Apparent Competition These examples reveal strong asymmetry in the effect of the shared natural enemy—the indirect interaction is mainly one-way. This is a common (not universal) feature of apparent competition. Explaining why asymmetries are often strong is still an open question and probably reflects both ecological factors and evolutionary history. In general, whichever species has the highest productivity (high r) and is not limited strongly by factors other than predation tends to dominate. If grey squirrels more effectively utilize a resource such as acorns, this could accentuate their dominance over reds in apparent competition. A full explanation of species extinctions from local communities often involves both competition (in the usual sense of the term) and apparent competition. These are not alternative, incompatible explanations but processes that can occur simultaneously and interact in various ways.

PROTECTING PREDATORS CAN PERMIT APPARENT COMPETITION TO OCCUR AMONG PREY

In the Kenyan example, humans historically suppressed zebras (which compete with livestock) and predators such as lions. Limiting livestock and reducing hunting in the interest of wildlife conservation allowed zebras to surge to high numbers, and predators such as lions reappeared. Predators do not substantially limit zebra numbers; zebras form large herds that provide protection from predation. But these herds do provide a steady supply of young, sick, or injured individuals that are easy pickings and can sustain predator populations. Other ungulates do not necessarily enjoy this kind of protection, and intensified predation appears to account for their declines.

Apparent Competition Does Not Arise Only in Disturbed Ecosystems In contrast to the invasion case studies, this African system involves species that have lived together for a very long time. Apparent competition is not just a process that shows up as a brief transient in

unnatural situations created by human disturbance and transplantation of species around the globe but can be important in natural ecosystems. The reason it was detected was that there was in effect a large-scale inadvertent experiment driven by a shift in land use patterns by humans. Stronger evidence for apparent competition in natural assemblages comes from deliberate experimental manipulations. In the rain forests of Belize, a rich community of leaf-mining insects (flies and beetles) sustains a high diversity of parasitoids. Experimental removal of some hosts led to lower parasitism in the remaining hosts, showing strong apparent competition in this natural community of herbivorous insects. Apparent competition could play a significant role in the dynamics of biodiversity over evolutionary timescales as well. Shared predation provides novel niche axes for specialization and diversification, and localized coadaptation of natural enemies and victims can lead to a sorting out of species among habitats or along gradients. Understanding the evolutionary dimensions of apparent competition is a largely unexplored area of theory.

Availability of Refuges Is a Key Element in Apparent Competition Another general message can be gleaned from the Kenyan example: the zebra has a partial refuge from predation by virtue of grouping behavior. Refuges come in many forms, ranging from permanent physical locations providing escape (e.g., rabbit warrens), to transient refuges in space (as in metapopulation dynamics), to escapes in time, to plastic adaptations that lower predation or parasitism rates. Stage structure (e.g., an invulnerable adult class) provides a particularly important form of refuge in some systems, and leads to rich complexities in theoretical models, because of the multiplicity of feedbacks that are possible (e.g., alternative stable states, and complex dynamics). If refuges protect some but not all individuals in a species, natural enemies can be sustained by this species without endangering it. Such species can dominate in apparent competition over species lacking refuges. Hosts can evolve to tolerate parasites or herbivores, without eliminating them. Such hosts could then exert strong apparent competition on alternative hosts that are not so well adapted. Understanding how population structure and evolutionary processes affect the strength of interspecific interactions, including apparent competition, is an active and growing area of ecological theory.

Multiple prey or host species can coexist, despite strong apparent competition, if each has its own refuge—a kind of niche partitioning in enemy-free space. An important subtlety is that this works if species are more likely to be in such a refuge when rare than when common. Theoretical models suggest that such refuge-mediated coexistence is greatly amplified if natural enemies behaviorally aggregate, spending more time where their victims are common.

PREDATOR BEHAVIOR GENERATES OTHER MECHANISMS OF APPARENT COMPETITION

Behavior—foraging tactics of predators and escape maneuvers of prey—is an important element in apparent competition. In the Channel Islands story, the presence of an abundant prey species, feral pigs, led to colonization by a few pairs of Golden Eagles, *Aquila chrysaestos.* Foxes on their own are far too scarce to warrant eagles setting up house on the island, but they are easy to casually catch as tasty morsels by eagles lured to the islands by an abundant alternative prey. This is called incidental predation in the literature.

In this case study, predator behavior is a key driver of apparent competition. The handful of golden eagles present could have nested on the mainland but instead chose to reside on the islands. In Fig. 2, rather than interpreting the isocline as describing the dynamics of an entire predator population (births and deaths), we can view it as a model of predator use of a small habitat patch (i.e., numbers change via immigration and emigration). When prey are scarce, predators leave; when prey are flush, predators arrive and stay. This aggregative numerical response is expected from optimal foraging theory; models predict that it leads to apparent competition between prey species within a patch, even if predator populations as a whole do not respond to shifts in prey numbers.

Flexible behavior (including patch use) can lead predators to ignore rare prey species, at least if those prey require special foraging tactics or specific recognition cues (search images), or if they are found in sites lacking other, more common prey. Many mechanistic processes lead to such switching by generalist predators. Theoretical models suggest that switching often promotes the persistence of multiprey assemblages. But counterexamples are also suggested by theory, dependent on the temporal scale and accuracy of switching relative to changes in prey abundance.

Changes in predator behavior that alter the indirect interaction between alternative prey exemplify what are called trait-mediated indirect interactions. There can be a range of effects alternative species have on each other, and the net effect is often context specific. One species can indirectly negatively impact another not by feeding a natural enemy but instead by in some other way facilitating

its presence or activity levels. Invasive shrubs can foster their own invasion by sheltering native herbivores from predators, so the herbivores more effectively reduce their native food plants, freeing resources for the invader. This is still apparent competition, albeit via a kind of ecological engineering. One direction for future theory will be to develop mechanistic models incorporating the multiplicity of potential channels of interactions among species.

A KIND OF APPARENT COMPETITION CAN OCCUR WITHIN INFECTED HOSTS

A major defense that vertebrates have against parasites is acquired immunity. To boil a complex story down to some basic elements, the immune system works by the stimulation of the proliferation of particular cell lines—a population of predatory cells—tailored in their attacks to parasites with certain attributes. If parasite A invades a host body and that host mounts an immune response, this may help fend off invasion by a relatively similar parasite B if it is recognized as being hostile by the host immune system. In this case, the natural enemy is the host immune system, and its numerical response (comparable to Eq. 1) is the growth of a particular population of defensive cells. The victims are different species or categories of parasites attempting to invade the host. Such apparent competition among similar strains of parasites is a powerful force selecting for parasite diversity.

This example illustrates the general point that abstract theoretical models can help illuminate comparable processes found in seemingly radically different systems. There are important differences between theoretical immunology and the theory of, say, vertebrate predator–prey interactions, but recognizing commonalities can help point to a unification of perspectives and approaches across levels of biological organization guided by powerful theoretical insights.

APPARENT COMPETITION IS ONLY PART OF THE STORY OF SHARED PREDATION

Does shared predation *always* lead to apparent competition? When it does, does it always tend to reduce prey species diversity? The short answer to both questions is "No!"

Let us go back to the basics. The simple theoretical model sketched above, part of which is shown in Figure 2A, makes many assumptions. Relaxing these assumptions can lead to a shift in theoretical predictions.

First Assumption We assumed an increase in prey numbers, for either of two prey species, benefits the predator (as assessed by a boost in its per capita growth rate). This might not hold, and for two reasons, one having to do with the prey themselves, and one having to do with the predator, largely independent of its prey.

Equation 1 assumes predator growth depends on prey availability. More generally, predator dynamics can depend on its own density. For instance, for successful reproduction weasels might require specialized nest sites, which could be in short supply and lead to direct competition. This could constrain the numerical response of weasels to mice and shrews, weakening apparent competition between these prey. Even if this is not the case, higher-order predators such as owls might limit weasel abundance. A pulse in mouse numbers might lead to a temporary increase in weasels, suppressing shrews, but in the long-term owl predation might bring the weasels back to their original levels. So food-web interactions can at times temper apparent competition or make it a transient response in system dynamics.

Equation 1 and Figure 2A assume consumption of each prey benefits the predator. This is not always true. Some prey contain toxins, harming consumers. Fish that die in the mass fish kills of red tides presumably are not very good at discriminating poisoned food from safe food. More subtly, even if a certain prey species is not absolutely bad, it might be bad in a relative sense, compared to other prey types. One generality about predators is that the rate at which they feed always saturates, due to limitations in handling time, gut capacity, or attention span. When food is abundant, time spent handling a low-quality prey type is in a sense wasted; there is an opportunity cost, because higher-quality prey are being ignored. Figure 2B shows how this alters the slope of the predator isocline. At sufficiently high abundance of the good prey, boosting numbers of the poor prey actually depresses predator growth; the indirect interaction shifts from $(-, -)$ to $(-, +)$.

This may seem unlikely, but something like this is believed to be of great importance in vector-borne infectious diseases involving multiple potential host species for the vector. Isoclines with partial positive slopes arise quite naturally in theoretical models of such systems. Some hosts are poor for pathogen reproduction, and if they can draw off attacks by vectors from more competent hosts, the net effect is a reduction in pathogen load across the entire system. Biodiversity can thereby moderate the risk of infectious disease.

Second Assumption We assumed interacting species reach equilibrium. Saturating responses, demographic stochasticity, and environmental fluctuations can lead to sustained oscillations. Theoretical studies show that alternative prey can sometimes boost each other's average numbers, when

one averages over the nonequilibrial dynamics of these highly nonlinear systems. Understanding how environmental variability and complex dynamics modify indirect interactions is a largely open area of theoretical ecology.

Third Assumption In our examples, apparent competition drove particular prey species to lower abundance or even extinction. Apparent competition via shared natural enemies can sometimes facilitate coexistence—if prey species also compete by interference or exploitatively for limiting resources. For instance, two prey species can persist on a single resource if a predator is present and one prey is better at resource competition and the other is better at apparent competition.

For a single generalist predator to enhance prey diversity beyond this effect requires some mechanism in effect permitting prey species to each have its own refuge from shared predation. As noted above, switching behavior by the predator is one such mechanism, as is habitat segregation among prey (given limited predator mobility). Another is to imagine that there is not just one predator species but instead a number, each to a degree specialized in its attacks to different prey. There can then be a kind of codependence of diversity across different trophic levels, and coexistence of numerous species in each trophic level becomes possible. Recent theoretical studies have explored this idea in some detail, highlighting the symmetry between shared predation and exploitative competition illustrated in Figure 1.

The symmetry is, however, incomplete. All species need resources, so there is always a potential for competition, within or among species. Whether or not shared predation leads to apparent competition is contingent (as we have seen) on many details. Moreover, there can be a difference in time scales over which dynamics play out. Resource levels can change very rapidly in response to changes in consumption (e.g., plants competing for light), whereas there can be significant time lags in responses of predators to their prey. Pathogen loads, however, can change rapidly, relative to host generation lengths, and for some purposes the impacts of shared pathogens and parasites might almost be viewed as a form of interference competition.

APPARENT COMPETITION PLAYS A ROLE IN HUMAN HISTORY AND OUR IMPACTS ON THE BIOSPHERE

Ecological theories such as apparent competition have implications for understanding ourselves and our impacts upon the rest of the biosphere. All the above case studies exemplify important problems in applied ecology. A clear appreciation of apparent competition theory can inform issues and management strategies in many practical disciplines, from conservation, to natural resource management, to pest control, to epidemiology, to invasion biology.

Apparent competition among peoples arguably has carved major channels in our own history. Parallel to the interaction between red and grey squirrels of the United Kingdom, plagues (in combination with other forms of more direct competition) have tragically influenced the waxing and waning of different peoples around the globe. For instance, as western Europeans conquered much of the world they had a hidden weapon—shared infectious diseases such as smallpox, to which indigenous populations were much more vulnerable. Further back in time, one hypothesis to explain the mass extinction of large mammals in North America is that it was overkill by a wave of colonizing people sweeping across the continent. Theoretical models suggest this hypothesis may work, if smaller species with a higher reproductive rate sustained the population of hunters, who could continue to preferentially attack large mammals such as mastodons even when the latter became vanishingly rare. In other words, apparent competition, mediated through humans, may be implicated in sculpting major features of the current fauna of entire continents.

Nearer the present, the ability of humans to hunt to extinction the Passenger Pigeon and Dodo reflects the fact that we do not depend on these species alone for sustenance. Humans are the ultimate generalist consumer, able to crack almost any victim's defenses, and our burgeoning population is sustained in terms of calories and nutrients by a quite small number of species, permitting us to wreak havoc upon the rest. The biodiversity crisis across the globe, seen through the lenses of ecological theory, may be an example of apparent competition, writ large.

SEE ALSO THE FOLLOWING ARTICLES

Conservation Biology / Ecosystem Engineers / Food Webs / Invasion Biology / Movement: From Individuals to Populations

FURTHER READING

Abrams, P. E. 2010. Implications of flexible foraging for interspecific interactions: lessons from simple models. *Functional Ecology* 24: 7–17.

Bonsall, M. B., and M. P. Hassell. 1999. Parasitoid-mediated effects: apparent competition and the persistence of host-parasitoid assemblages. *Researches on Population Ecology* 41: 59–68.

Chesson, P., and J. J. Kuang. 2008. The interaction between predation and competition. *Nature* 456: 235–238.

DeCesare, N. J., M. Hebblewhite, H. S. Robinson, and M. Musiani. 2010. Endangered, apparently: the role of apparent competition in endangered species conservation. *Animal Conservation* 13: 353–362.

Harmon, J. P., and D. A. Andow. 2004. Indirect effects between shared prey: predictions for biological control. *BioControl* 49: 605–626.

Holt, R. D., and J. H. Lawton. 1994. The ecological consequences of shared natural enemies. *Annual Review of Ecology and Systematics* 25: 495–520.

Mideo, N. 2009. Parasite adaptations to within-host competition. *Trends in Parasitology* 25: 261–269.

Oliver, M., J. J. Luque-Larena, and X. Lambin. 2009. Do rabbits eat voles? Apparent competition, habitat heterogeneity and large-scale coexistence under mink predation. *Ecology Letters* 12: 1201–1209.

APPLIED ECOLOGY

CLEO BERTELSMEIER, ELSA BONNAUD, STEPHEN GREGORY, AND FRANCK COURCHAMP

University of Paris XI, Orsay, France

Applied ecology aims to relate ecological concepts, theories, principles, models, and methods to the solving of environmental problems, including the management of natural resources, such as land, energy, food or biodiversity.

DEFINITION AND SCOPE

What Is Applied Ecology?

Despite its somewhat restrictive name, applied ecology is more than simply the application of fundamental ecology. In a nutshell, ecological management requires prediction, and prediction requires theory. Applied ecology is a scientific field that studies how concepts, theories, models, or methods of fundamental ecology can be applied to solve environmental problems. It strives to find practical solutions to these problems by comparing plausible options and determining, in the widest sense, the best management options.

One particular feature of applied ecology is that it uses an ecological approach to help solve questions concerned with specific parts of the environment, i.e., it considers a whole system and aims to account for all its inputs, outputs, and connections. Of course, accounting for everything is no more possible in applied ecology than it is in fundamental ecology, but the ecosystem approach of applied ecology is both one of its characteristics and one of its strengths.

Indeed, one could view the overall objective of applied ecology as to maintain the focal system while altering either some of the elements we take from the system (i.e., ecosystem services or exploitable resources) or some of those we add to the system (i.e., exploitation regimes or conservation measures) through an educated management strategy. Since those two types of elements are not mutually independent, long-term management strategies are best aimed at optimizing rather than maximizing exploited items. This is more efficiently achieved through an adequate understanding of theoretical ecology, which generally considers all parts of the system rather than a limited set of its components.

What Are the Fields Covered?

The word "applied" implies, directly or indirectly, human use or management of the environment and of its resources, either to preserve or restore them or to exploit them. Humans influence the Earth at all levels: the atmosphere, the hydrosphere (oceans and fresh water), the lithosphere (soil, land, and habitat), and the biosphere. Understandably, questions related to human populations (notably its demography) fall within the scope of applied ecology, as most impacts on ecosystems are directly or indirectly anthropogenic.

Aspects of applied ecology can be separated into two broad study categories: the outputs and the inputs (Fig. 1). The first contains all fields dealing with the use and management of the environment for its ecosystem services and exploitable resources. These can be very diverse and include energy (fossil fuel or renewable energies), water, or soil. They can also be biological resources—for their exploitation—from fish to forests, to pastures and farmland. They might also, on the contrary, be species we wish to control: agricultural pests and weeds, alien invasive species, pollutants, parasites, and diseases. Finally, they can be species and spaces we wish to protect or to restore.

The fields devoted to studying the outputs of applied ecology include agro-ecosystem management, rangeland management, wildlife management (including game), landscape use (including development planning of rural, woodland, urban, and peri-urban regions), disturbance management (including fires and floods), environmental engineering, environmental design, aquatic resources management (including fisheries), forest management, and so on. This category also includes the use of ecological knowledge to control unwanted species: biological invasions, management of pests and weeds (including biological control), and epidemiology.

The inputs to an applied ecology problem consist of any management strategies or human influences on the target ecosystem or its biodiversity. These include conservation biology, ecosystem restoration, protected area design and management, global change, ecotoxicology

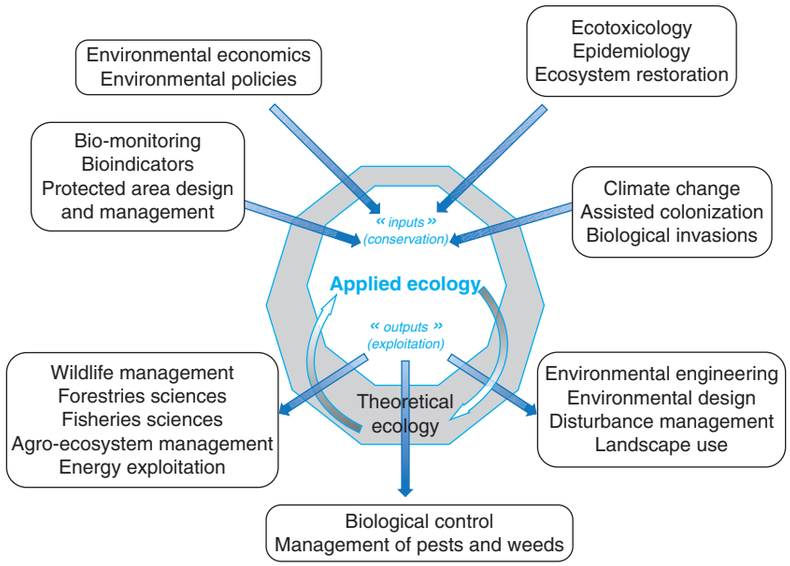


FIGURE 1 The relationships between the different fields of applied and theoretical ecology.

and environmental pollution, bio-monitoring and bio-indicators of environmental quality and biodiversity, environmental policies, and economics. Of course, these outputs and inputs are intimately connected. For example, the management of alien invasive species is relevant to both natural resource management (e.g., agriculture) and the protection of biodiversity (conservation/restoration biology).

In addition to using fundamental ecology to help solve practical environmental problems, applied ecology also aspires to facilitate resolutions by nonecologists, through a privileged dialogue with specialists of agriculture, engineering, education, law, policy, public health, rural and urban planning, natural resources management, and other disciplines for which the environment is a central axiom. Indeed, some of these disciplines are so influential on environmental management that they are viewed as inextricably interlinked. For example, conservation biology should really be named "conservation sciences" because it encompasses fields that are not very biological, such as environmental law, economics, administration and policy, philosophy and ethics, resources management, psychology, sociology, biotechnologies, and more generally, applied mathematics, physics, and chemistry.

And obviously, as we are dealing with environment and ecosystems, everything is connected, all questions are interrelated, all disciplines are linked, and all answers are interwoven. Understanding a process through one field of applied ecology will allow advancing knowledge in other fields.

What Are the Links between Applied and Theoretical Ecology?

Because theoretical ecology may be defined as the use of models (in the widest sense) to explain patterns, suggest experiments, or make predictions in ecology, it is easy to see that relations between theoretical and applied ecology are bidirectional. Simply put, theory feeds application, but application also allows for the testing of theory. Indeed, applied ecological problems are used to assess and develop ecological theory. In this regard, these two aspects of ecology complement and stimulate each other.

Yet the links between theoretical and applied ecology can fray. Theoretical ecology operates within the bounds of plausibility. What is theoretically plausible, however, is not always ecologically realistic. And if it is not ecologically realistic, then theoretical ecology cannot be applied to interrogate real ecological situations. In other words, theoretical ecology cannot always be used in applied ecology. The links between theoretical and applied ecology range from spurious to robust and have been used with varying success in different fields. Fields that have benefited from theory include fisheries and forestry management and veterinary sciences and epidemiology (both human and nonhuman). However, some other fields of applied ecology have not (or not yet) benefited fully from ecological theories, concepts, and principles, and that aspect is the focus of this entry.

Three such examples are presented here. The first is an illustration of a major field of ecology that has a strong applied branch but has not, until very recently, fully

taken advantage of theoretical ecology: invasion biology. The second describes a theoretical process important to many aspects of applied ecology (including epidemiology, fisheries, biological control, conservation biology, and biological invasions) but which has nonetheless been underused so far in applied ecology: the Allee effect. The third depicts an emerging field that is now, by obligation, mostly theoretical but which has an applied future and for which it is hoped that ecological applications will emerge rapidly: climate change.

APPLIED ECOLOGY WITHOUT THEORETICAL ECOLOGY: BIOLOGICAL INVASIONS

The study of biological invasions is one striking case where applied ecology has historically been at the core of the discipline's development without fully benefiting from theoretical ecology. Alien invasive species are those species introduced beyond their ordinary geographical range by human agency and that are infamous for causing ecological and economical damage. Globally, they are considered as the second biggest cause of species extinctions after habitat degradation, but they reign as the single most devastating impact in island ecosystems.

The seventeenth century marked the start of the vigorous and uninterrupted human colonization of the world's islands, initially by sailors, pioneers, settlers, hunters and fishermen, and more recently by militaries and tourists. Whether by a poor understanding of ecosystem functioning or a scarce consideration for the environment, this colonization led to many ecological disturbances, in particular the introduction—deliberate or accidental—of many exotic plant and animal species. Although not confined to islands (exotic organisms have been introduced to continents, lakes, and even oceans), the introduction of exotic organisms to islands has often led to dramatic changes in their communities. This is in part because insular ecosystems are fragile and characterized by simple trophic webs, vacant ecological niches, and low species diversity. Perhaps more important, however, is the fact that native species often evolve in the absence of strong competition, herbivory, parasitism, or predation and are less adapted to these newcomers. Moreover, islands harbor high levels of endemism, several times higher than comparable continents ecosystems, which makes every population extirpation a probable species extinction. As a result, alien invasive species on islands have affected countless island plant and animal communities and have been responsible for numerous native species extirpations and extinctions (nearly 60% of modern species extinctions have occurred on islands).

The first attempts at controlling alien invasive species included mechanical (trapping, hunting) and chemical (poisoning) methods, but the majority were biological: deliberate introductions of the alien invasive species' natural enemies to efficiently find and kill them. For example, numerous attempts at controlling invasive rodents have involved releasing cats, dogs, or mongooses. Prior to many of these attempts, it had not been realized that even the most efficient predators very seldom remove an entire prey population but instead control it at reduced numbers. This is because invasive rodents have evolved with cat predation and have developed antipredator behaviors and high reproductive rates. Unfortunately for the native species that evolved in the absence of these predators and whose populations were impacted by them, this problem could have been anticipated since it is a fundamental prediction of predator–prey models and evolutionary ecology theory.

A famous example of such misguided attempts is the case of the repeated endeavors at controlling rodents in sugarcane fields in Jamaica during the nineteenth century. There, cane growers introduced ants (*Formica omnivora*), which did not reduce rat numbers but soon became a problem themselves. To remove rats and ants together, it was then decided to introduce toads (*Bufo marinus*). But toads still did not control rats, and they too became a pest themselves. Finally, small Indian mongooses were introduced to control rats and toads. Mongooses failed to control either and began preying on native birds, posing new threats to wildlife.

Another illustration of using biological methods in the absence of a careful theoretical framework is the use of pathogens to control invasive rabbits. The introduction of the myxoma virus into both Europe and Australia to reduce rabbit numbers, and the more recent introduction of rabbit haemorrhagic disease (RHD) into New Zealand for the same goal, were both designed without a proper ecological framework and performed neither by ecologists nor by conservation managers. These two cases point out how difficult it may be to fully control the intentional introduction of microorganisms by persons who are not fully aware of (and/or interested in) the potential ecological effects of such actions. Theoretical epidemiology was, however, sufficiently advanced to predict these unfortunate outcomes. Sadly, history now shows that the absence of a rigorous theoretical framework often renders these empirical attempts to apply simple ecological principles (e.g., predator–prey or host–parasite interactions) to real ecological problems prone to failure.

Unfortunately, early conservation managers, too, used trial and error to design island restoration projects, to the

point that the whole subdiscipline of alien invasive species control has been long dominated by empirical development of techniques and methods at the expense of the use of ecological theory. For example, chemical advances have led to powerful, specific poisons, and ingenious trapping devices have outwitted even the most cunning invasive mammals. But these are technological developments that do not rely much on theory. With the notable exception of Virus-Vectored immuno-Contraception, in which empirical and theoretical studies were developed in parallel, most of the progress made in this field has neglected the potential benefits of theoretical ecology. This understandably might have been because there was an urgent need for a specific application. Yet this general trend has also led to many delays (and even some failures) in eradication programs.

Recent advances to address failing eradication programs have spawned adaptive control programs. This second generation of control programs capitalizes on theory by using ecological concepts from population dynamics and behavioral ecology. Such theory has been used in a number of ways. For example, it has advised that the timing or order of species eradications can be used maximize their success, that techniques such as the "Judas goat" technique can exploit herding behavior to locate the last individuals, or that helicopters and GIS can be used to deliver baits more accurately. More recently, it has allowed us to develop a better understanding of spatial ecology in the first stages of rodent invasions on islands, in order to better detect and control them at this crucial step.

Despite the improvements in the second-generation adaptive control programs, they still often lack an ecosystem perspective, leading to a general underappreciation of the importance of chain reactions following sudden alien invasive species eradications. As, understandably, nongovernmental organizations, conservation managers, and wildlife bodies have to react faster than the rate at which fundamental research can operate, some eradication plans have suffered from a lack of pre-eradications studies aimed at understanding the direct and indirect biotic interactions linking native and alien invasive species. As a result, some exotic species that were held in low densities by alien invasive species exploded once the pressure from the alien invasive species was removed following their eradication, leading to further damage to the native ecosystem. For example, invasive goats were recently removed from the threatened native forest on Sarigan Island, in the tropical western Pacific Ocean. Unfortunately, goats were selectively suppressing an exotic vine (*Operculina ventricosa*), which was highly competitive compared to other native plants and eventually exploded following the release from goat browsing.

Although a few empirical examples exist, the study of these "surprise" secondary effects of alien invasive species eradications has been largely theoretical. Take, for example, the mesopredator release effect. In theory, when a native species is a shared prey of introduced predators (e.g., cats and rats), the removal of the introduced top predator (cat) might result in a numerical increase of the introduced mesopredator (rat) population. This increase may be highly detrimental for the native species. In this case, these biotic interactions were studied theoretically through a mathematical model, and there have been few empirical studies of this "surprise" effect. This situation seems to occur only under certain ecological circumstances (e.g., absence of alternative food resources) and depends on which species are affected. Similarly, the competitor release effect suggests that when controlling for a higher competitor (e.g., a rat), the population of the lower competitor (e.g., a mouse) may erupt, even if these are controlled at the same time. This is simply because under certain conditions the direct effect of the removal is lower than the indirect, positive effect of the competitor's removal. In that case, the more severe the control, the more sudden the competitor release. This may explain why in several cases the eradication of rats was followed by a dramatic explosion of hitherto overlooked introduced mice.

Ideally, eradication programs should be based on a thorough knowledge of the interactions between the species involved, in particular regarding other introduced species. This principle has been applied to a long-term study that included removal invasive black rats from Surprise Island in the Entrecasteaux Reef, New Caledonia. Over a 4-year period, the island flora and fauna was characterized qualitatively and quantitatively to study the impact of rats on native species but also to reveal the presence of other introduced species: mice, ants, and plants. A thorough study of the trophic relationships within the invaded ecosystem, both empirical and theoretical (including mathematical modeling), led to the design of a case-specific eradication strategy that enabled removal of the rats without triggering a release of other invaders (e.g., simultaneous mouse removal was planned into the case-specific protocol). Then, this ecosystem has been the subject of a long-term post-eradication survey to follow both the recovery of the local communities and the potential emergence of surprise effects following the removal of rodents. This shows how theoretical ecology (here trophic ecology) may be helpful in imporving the efficiency of applied ecology programs.

Because it is not possible to conduct long pre-eradication studies before each eradication control, as was done for Surprise Island, the establishment of robust control programs rooted deeply in theoretical ecology will be necessary to achieve the full potential of this central area of applied ecology.

THEORETICAL ECOLOGY WITHOUT APPLIED ECOLOGY: THE ALLEE EFFECT

In contrast to the case of biological invasions, population dynamics, and in particular the domains developing around the Allee effect, is a discipline deeply rooted in theoretical ecology, but it still lacks the full transfer to applied ecology.

A demographic Allee effect describes a reduced per capita population growth rate in a population reduced in size (or density). Theory suggests that many, or even most, species could be susceptible to a demographic Allee effect, either directly, or indirectly through interaction with another species. As a consequence of this perceived ubiquity, it has become a cornerstone concept of ecology, with many potential applications to population management. Since its origins in the 1940s, the Allee effect has been the subject of roughly two types of studies: those aiming at a better understanding of the mechanisms underpinning it and its theoretical implications, and those trying to assess the presence of this process in given taxonomic groups. Probably because the concept is intellectually exciting, but also because empirical and experimental approaches are generally more difficult, the theoretical line of study has been the most active.

Another reason for the dominance of theory-based studies in this field is that empirical demonstrations of Allee effects are scarce and have mostly been restricted to mechanisms affecting individual fitness (component Allee effects) rather than manifestations in population dynamics (demographic Allee effects). There are sound biological reasons why demographic Allee effects might be the exception rather than the rule, although these are irrelevant to the point made here. What matters is that the majority of Allee effect studies, particularly over the last decade, have been theoretical studies. Consequently, there is now a thorough understanding of the Allee effect based on its conditions, mechanisms, and implications in a variety of contexts, including single populations and metapopulations, interspecific interactions, and spatially explicit frameworks. The importance of this process for different fields of applied ecology, including conservation biology, species management, and population exploitation, is also well understood. Yet studies on Allee effects conducted in the realm of applied ecology remain unfortunately scarce.

The fact is that managers and conservationists should be concerned about Allee effects, even in populations that are not apparently exhibiting them, because (i) they can create critical thresholds in size or density below which a population will crash to extinction, and (ii) even if there is no threshold, they increase extinction probability (due to stochasticity) and reduce per capita population growth rate in low-density or small populations. Whether extinction is a desirable or dreaded outcome, such knowledge may be critical for effective population management. For example, field studies of the spread of the wind-pollinated invasive cordgrass *Spatina alterniflora* in Willapa Bay, USA, reveal that it has been slowed by the existence of an Allee effect affecting plants in low-density areas, such as those that characterize the leading edge of the invasion, and this information has been central to control strategies.

Clearly, there is an important need for a whole line of work aimed at applying our theoretical understanding of Allee effects to solve real practical problems, particularly in those areas for which theory suggests it will have its greatest influence. One pressing situation that would deeply benefit from this transfer is the study of the conditions needed to maintain the population above the extinction threshold. This task will not be easy, especially since the extinction threshold is difficult to estimate and will vary according to both species and circumstance—habitat quality, mortality rates from predator or exploitation, and so on. But the benefits would likely be worth the effort. For example, it might be applied to invasion biology to enhance the efficiency of alien invasive species control programs. Theory states that targeting every individual of an introduced population might be unnecessary if the species exhibits a demographic Allee effect, with a substantial financial saving to be gained from leaving the last individuals to die out by themselves. Similarly, Allee effect theory can be used for the control of insect pests in agroscience. Several studies have investigated minimum propagule sizes for successful establishment of agents used in biological control programs.

Apart from management of introduced populations, Allee effect theory can also be used, somewhat obviously, in the management of threatened populations, affording managers an early warning system to avert their imminent extinction. For example, Allee effects interact synergistically with mortality from exploitation, which not only reduces the population size but can also create

or increase an extinction threshold. Exploitation of populations with possible Allee effects needs to be considered and managed very carefully. Examples of fishery collapses (and lack of recovery) show how the neglect of Allee effect theory in the implementation of maximum sustainable yields and quotas resulted in disaster. In 2006, Courchamp and colleagues showed that exploitation itself can also act as a mechanism to create an Allee effect, particularly if the economic value of a species is found to be increasing with rarity, and the implications of this new process remain to be studied. In addition to trying to preserve small and decreasing populations, it might be better to augment their size or distribution via translocation or population reinforcement. Just as Allee effects affect alien invasive species eradication campaigns, they will also alter reintroduction strategies. For example, reintroductions of quokkas and black-footed wallabies to Western Australia were met with failure because the sizes of the populations released were too small to overcome the sustained predation pressure from introduced cats and foxes.

Even studies of the second type, those aiming at finding an Allee effect in specific taxonomic groups, have suffered from a failure to use theory. It is now common knowledge that Allee effects may occur in a wide number of species and that the presence of this process may dramatically alter the conditions and chances of population viability. Yet entire taxonomic groups of conservation interest, including, for example, bats, primates, and felids, remain to be investigated in this context. Whether for conservation purposes or in relation to an economic interest, eusocial insects should also be a focus. For example, bees are of commercial importance and ants are of major relevance as biological invaders, and both are major providers of crucial ecosystem services. Yet these and the other tens of thousands of eusocial species of ecological importance have so far not been the subject of Allee effect studies (Fig. 2). Perhaps the only example of such a subject for an Allee effect study is a colonial spider (*Anelosimus eximius*), whose lifetime reproductive success (a composite measure of offspring production and survival) increases with colony size.

In many contexts, such as the study of alien invasive species, applied ecology has advanced thorough empirical studies despite the existence of a wealth of relevant theory. In other contexts, such as the study of Allee effects, theoretical ecology has been at the forefront of our understanding. Now, the time is right to use our theoretical understanding of processes such as the Allee effect to benefit applied ecology.

FIGURE 2 Argentine ants (*Linepithema humile*), as an illustration of a species fitting the three examples detailed in this entry. They are one of the worst biological invaders in several parts of the world. In southern Europe, this species forms a supercolony from Spain to Italy and is likely to present an Allee effect. There, it seems unable to expand its range northward of its current area of invasion because of unsuitable climatic conditions at higher latitudes, but models predict that this is likely to change with global climate change. Photograph by Alex Wild.

THEORETICAL ECOLOGY FOR APPLIED ECOLOGY: CLIMATE CHANGE

A final example concerns an emerging area that has, up to now, mostly benefited from theoretical ecology, but which promises to be of major importance for future developments in applied ecology: climate change.

Over the last two centuries, anthropogenic activities have caused rapid climate change, with the mean global annual temperature expected to rise by 1–3.5 °C by 2050. Although climate change encompasses a wide range of phenomena like changes in precipitation regimes, increases in extreme event frequency, ocean acidification, and sea level rise, the vast majority of studies concern rises in temperature, which currently figures among the biggest threats to biodiversity. Species' responses to climate change have already been observed in many taxa. They include physiological changes, phenological changes (i.e., the timing of life cycle events such as plant flowering and fruiting or animal migration and reproduction), and geographic range shifts where species migrate to an area with a more favorable climate. If a species fails to adapt or to shift their ranges in response to climate change, then they may go extinct. This is the major prediction of many studies that have projected global biodiversity loss due to climate change. In one of the most widely cited studies, Chris Thomas and colleagues estimate that 15–37% of species are "committed" to extinction. It is clear that there is an urgent need to revise conservation strategies in

the face of global climate change, and because these predictions are based on events yet to pass, we must turn to the theory of climate change science to serve as the basis of these revisions. Because it is based on projections, the field of climate change biology has taken off by mainly developing a foundation of mathematical models that aims to predict species' potential ranges and fates under different scenarios.

There are two main categories of predictive models commonly used in climate change science. The first category represents the most widely used bioclimatic envelope models. They relate a current species' range to multiple environmental variables (such as temperature, precipitation, and so on) and thereby define the climatic niche of that species (its potential spatial distribution based on these variables), making the assumption that it reflects its ideal climatic conditions. Under different climate scenarios, the species' predicted future geographical distribution can be identified. Species whose predicted distribution will be no longer be within their climatic envelope are predicted to go extinct.

The second category of models is based on a species' physiology. For example, dynamic global vegetation models estimate plant biodiversity loss following climate-induced changes in geochemical cycles and CO_2 concentrations because of their physiological constraints. Physiology-based models that can be applied to animals include degree-day models, which estimate the number of days per year that the temperature is above a critical minimum temperature—the temperature below which the species cannot survive. These models are particularly useful for ectotherms.

These models can, of course, be complemented to some extent by experimental studies (based on increased temperature) or comparison with past data (some effects of global warming have been observable for the last several decades), but for now the major source of ecological knowledge comes from modeling. Yet the major interest for ecology in this domain will probably be in applied ecology. For example, bioclimatic models can help to reevaluate the current set of protected areas. They can help to identify climatic refuges and heterogeneous microclimatic conditions that may be important to save a species from extinction. Furthermore, models can help to prioritize protected regions. For example, islands that are less susceptible to sea-level rise should be favored both for protection and as refuges for threatened species. A widespread view is that an important strategy is to enhance landscape connectivity to enable species to move through a matrix of interconnected habitats in order to escape from unsuitable climatic conditions. However, as haunts the previous debates on corridors, enhancing landscape connectivity can also be argued as counterproductive because populations in the lagging edge of their advancing range would compete with arriving populations and communities. On the other hand, species range shifts might be the unavoidable consequence of climate change because a species' existing range might not overlap with its predicted future range. As this situation could lead to extinction, climate change has been a major argument for the proponents of human-assisted colonization. In this increasingly hot debate, it is likely that the solution might come from a solid blend of robust theoretical ecology and educated case-by-case decisions based on the known ecological characteristics of the concerned species and the socioeconomic context in which conservation is taking place. Surely theory has a major role to play here, too.

Bioclimatic models can also be used to explore the impact of different drivers of global change on biodiversity. For instance, it is very important to get a better understanding of how alien invasive species will react to climate change and whether they might benefit from shifting climatic envelopes to invade new areas previously outside their distribution. Models can help to identify species posing the greatest risk for invasion of a particular region under a scenario of climate change (Fig. 3).

Moreover, climatic models can also serve to evaluate the risks of disease spread following a rise in temperature, for example. Many disease vectors are very sensitive to climatic conditions, and there is a general fear that tropical diseases might extend to higher latitudes, leading to a dramatic increase of infectious diseases, especially when local hosts are naïve to the new parasites. This concerns plant, wild animal, domestic animal, and human populations.

These theoretical studies can help to estimate risks to biodiversity (invasions, diseases, species losses) and to prioritize protected areas. However, as all these projections of future biodiversity are based on models, they include the uncertainties inherent in their input variables, including projected climate change and the type of species' responses. They can thus be validated only indirectly, by extrapolating from observed changes over the last decades and through independent data sets of current species' occurrences. Even then, these models do not account for possible behavioral, phenological, or physiological adaptations and do not adopt an ecosystem perspective (communities will react differently than the sum of species

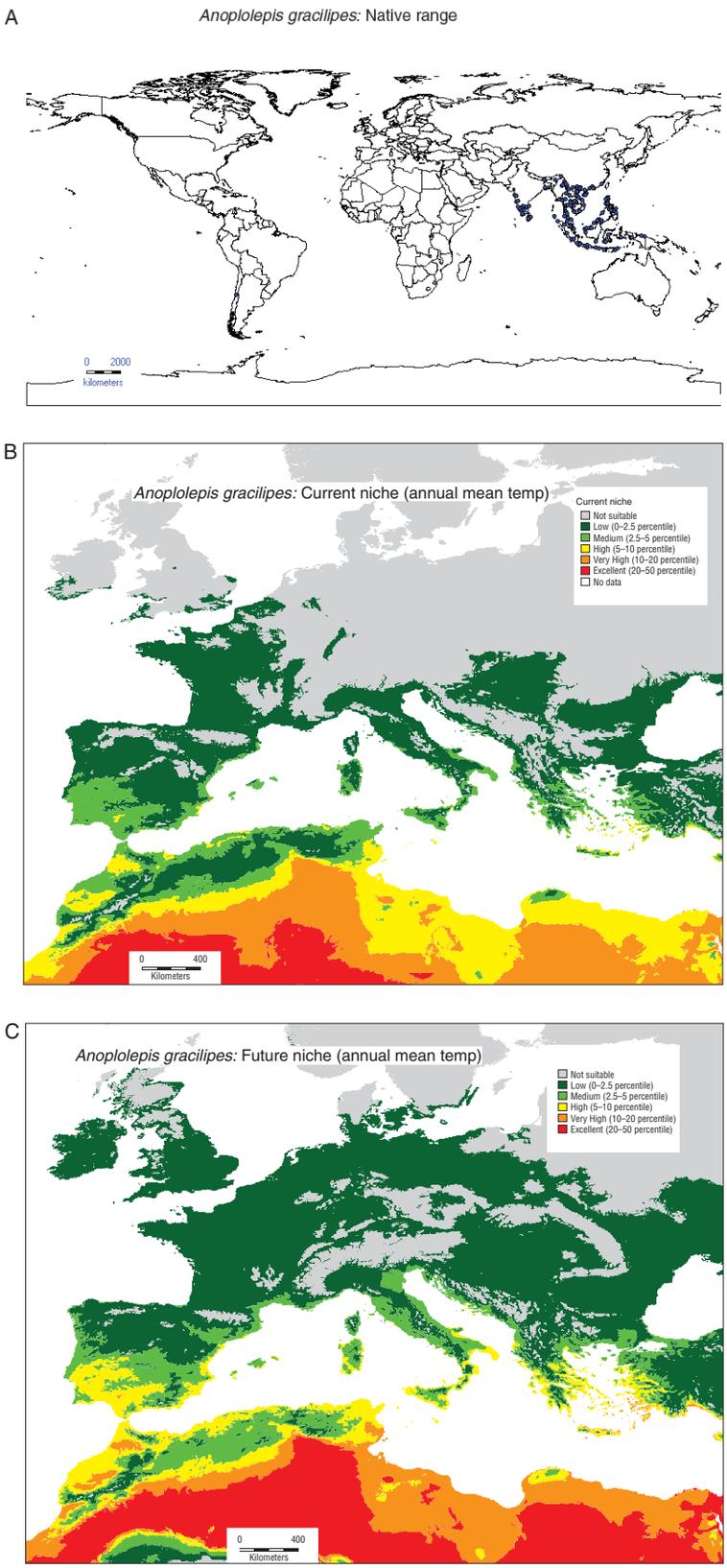


FIGURE 3 Bioclimatic envelope models for the yellow crazy ant (*Anoplolepis gracilipes*) that relate the species' native distribution (A) to climatic variables (e.g., temperature) in order to determine the species' current climatic range that can then be projected on a climate map to determine its potential range based on climatic conditions in 2010 (B) and under a scenario of climate change, for example in 2050 (C).

reactions). Paradoxically, one may claim that the current theoretical approaches should be more closely based on theoretical ecology. Indeed, there is an urgent need to improve model predictions because many future conservation decisions will have to be made rapidly with the help of climate change model predictions.

Climate change science is, by its very nature, a predictive science, using today's concepts and models in developing a robust theoretical understanding of future climate change to advance the applied ecology of tomorrow.

CONCLUSION: WHAT NOW?

This entry deals with the antithesis of theoretical ecology—applied ecology. A vast number of disciplines fall within its realm, and applied ecology provides theoretical ecology with a *raison d'être.* Applied ecology can (and should) benefit more from theoretical ecology, either by investigating new fields of applied ecology or by working to apply new concepts of theoretical ecology. This means not only that theoretical ecologists can look forward to a whole new set of questions, approaches, and tools to study, but also that they will have new subjects to explore and new colleagues to interact with.

The relationships between theoretical and applied ecology are better defined by a feedback loop than by a simple, unidirectional supply link. Enhancing the connections between these two complementary aspects of ecology will not only help solve pressing environmental issues but will also further stimulate theoretical ecology.

SEE ALSO THE FOLLOWING ARTICLES

Allee Effects / Conservation Biology / Demography / Ecosystem Ecology / Fisheries Ecology / Invasion Biology / Restoration Ecology

FURTHER READING

Cadotte, M. W., S. M. McMahon, and T. Fukami, eds. 2006. *Conceptual ecology and invasion biology.* Berlin: Springer.

Caut, S., E. Angulo, and F. Courchamp. 2009. Avoiding surprise effects on Surprise Island: alien species control in a multi-trophic level perspective. *Biological Invasions* 11(7): 1689–1703.

Courchamp, F., J. Gascogne, and L. Berek. 2008. *Allee effects in ecology and conservation.* Oxford: Oxford University Press.

Courchamp, F., E. Angulo, P. Rivalan, R. Hall, L. Signoret, L. Bull, and Y. Meinard. 2006. Value of rarity and species extinction: the anthropogenic Allee effect. *PLoS Biology* 4(12): e415.

Lovejoy, T., and L. Hannah. 2005. *Climate change and biodiversity.* New Haven: Yale University Press.

Newman, E. I. 2001. *Applied ecology & environmental management.* London: Blackwell Science.

Shigesada, N., and K. Kawasaki. 1997. *Biological invasions: theory and practice.* Oxford Series in Ecology and Evolution. Oxford: Oxford University Press.

Thomas, C. D., A. Cameron, et al. 2004. Extinction risk from climate change. *Nature* 427(6970): 145–148.

Williamson M. 1997. *Biological invasions.* Population and Community Biology Series. Berlin: Springer.

ASSEMBLY PROCESSES

JAMES A. DRAKE AND PAUL STAELENS

University of Tennessee, Knoxville

DANIEL WIECZYNSKI

Yale University, New Haven, Connecticut

Ecological systems are dynamically nonlinear, self-organizing, spatially extended, and multistable, driven by species colonization and extinction events, all operating against a backdrop of environmental and spatial variation, disturbance, noise, and chance. System assembly or construction is fundamentally important to understanding ecological systems.

ECOLOGICAL REALITIES

The nature of ecological systems creates a number of epistemological difficulties for the observer. When historical events have shaped the current system state, it becomes difficult for the observer to assign cause without knowledge of such events. For example, variation in the order of arrival and timing of species colonization has been shown to generate alternative community states. In fact, Drake, in 1990, found that assembling communities not only result in irrevocable differences in structure, but if perchance the developing trajectories cross (e.g., identical food webs), at some time they respond differently to a common invader. Such behavior is a consequence of both where each system came from and where they are going. This historical disconnect creates a very real danger when fundamental principles or cause are assigned based on our necessarily brief analyses of presently operating mechanisms, processes, and cycles. It is also unclear whether any given system is riding along a transient heading elsewhere or has reached some solution. This limits the power of experimentation to temporal slices of assembly time. How can the observer know whether the mechanisms and processes experimentally documented in a system are causal or are simply those permitted at a given time along some trajectory of development? Understanding these permissions is key to understanding biological nature. But how to proceed?

Manipulating system development from its inception to some end state is an approach to understanding ecological systems termed assembly. Assembly experimentation and theory seek to describe the creation of novel attractors that accompany colonization and extinction. This approach builds dynamics and explores the structural consequences of traversing alternative dynamical realms. For example, simply by comparing systems created by permuting the order of species arrival, the multistable character of ecological nature has been exposed. By analogy, consider a simple jigsaw puzzle. Here, the arbitrary placement of any given piece at any time does not alter the final outcome. There is but a single solution. However, should the puzzle pieces exhibit dynamics of their own, as is the case with species populations, multiple solutions arise and a finite set of parts produces multiple images.

Assembly processes are the events and subsequent dynamical response that arise during the construction, succession, and self-organization of a system. These events initiate and drive the trajectory of a system by redefining the systems attractor, its basin, and the behavior of the system along alternative transients. Therefore, assembly processes generate a map of species colonization, interactions, extinction, and the development of a topology or network (e.g., food web). It is important to note that this transient stage of development can be very long lived. An ecological attractor might be of a simple sort, the system arriving in uneventful fashion to some stationary state (Fig. 1). Just as likely, however, an attractor may itself contain complex dynamics being composed of a series of dynamically varies states. Here, for example, the system proceeds through a period of regular dynamics, followed by period doubling and chaos, and residency on a strange attractor.

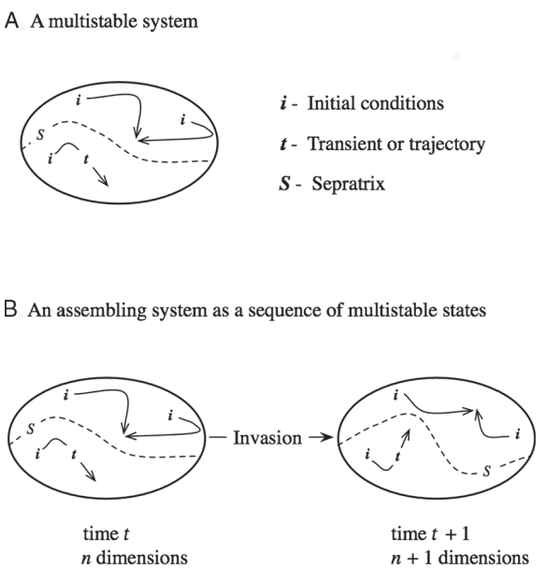


FIGURE 1 System trajectories in a multistable landscape (A), and a higher-order trajectory (B) created by increasing system dimensionality.

The assembly approach exploits the fact that complex systems exhibit sensitive dependence to initial conditions. Here, arbitrarily small changes in initial population sizes can create novel trajectories to a systems solution. Sensitive dependence creates a situation where a given system, variously initialized, experiences regions of state space unique to particular sets of initial conditions. Consequently, system properties and the action of mechanisms may be fundamentally different as a function of history. Further, ecological systems are open in the sense that they can be colonized. System dimensionality and degrees of freedom are not fixed but vary with time and over space. For example, a species might successfully colonize a system, but fail to invade that system if the system developed from a slightly different set of initial conditions. Because species composition varies over the course of community assembly, ecological systems experience a series of initial conditions and subsequent restructuring of dynamical character as a function of dimensionality changes.

PATTERN AND ANTECEDENT CAUSE

The idea that historical influences play a strong if not defining role in the determining species composition of a community is hardly new. Henry Cowles, for example, wrote in 1901:

> Antecedent and subsequent vegetation work together toward the common end. Where there is no antecedent vegetation, *Ammophila* and other herbs appear first, and then a dense shrub growth. . . .

Cowles had surely observed sequence and priority effects, both hallmarks of community assembly. Even earlier, Dureau de la Malle (1825) visualized alternative themes in forest regeneration and suggested that such themes could reflect fundamental natural laws. Despite these early conceptual advances, the temporal scale of community assembly and succession has hindered direct experimentation in the real world.

Inferential approaches to revealing historical effects are based on the assumption that recurrent patterns are a consequence of dynamical commonalities among developing or assembling communities. Both Elton (in 1946) and Hutchinson (in 1959) sought explanations for observed patterns in the action of mechanisms like competition occurring in the past. The allied ideas that as community development proceeds rules are invoked

tuning invasion success around a particular species-to-genus ratio and patterning in gape size belying limits to similarity are strikingly elegant yet difficult to document in nature. This is as it should be because system assembly is occurring in an environment subject to disturbances of various magnitudes and at multiple scales. Should disturbance ameliorate competitive interactions, system development is reset or comes under the control of a new attractor, and vulnerability to invasion is modified. Of course, it is also possible that system dynamics and the species that generate those dynamics have variously incorporated disturbance as a resource.

In 1975, Jared Diamond took this approach further by suggesting that presently observed patterns among a set of communities could be used to infer dynamical rules that govern system development. Such rules are ostensibly a higher-level summary of the mathematical intricacies that underly species interactions and topological development in space and time. For example, Diamond observed that particular species pairs never coexist, a phenomenon seductively titled forbidden combinations. When species in some forbidden combination exhibit priority effects, where the first species to colonize wins, an assembly rule emerges. While Diamond's arguments were met with fervent resistance on statistical grounds, simple null models were capable of generating similar patterns; his discourse stimulated a vigorous and ongoing experimental and theoretical effort.

A more direct approach was taken by Piechnik, Lawler, and Martinez in 2008. They reexamined Simberloff and Wilson's arthropod recolonization data and found that trophic breadth was reflected in species colonization order during community assembly. Early invaders functioned as generalists, while successful colonists late during community development were specialists. This result could fit nicely with Petermann and colleagues' 2010 study implicating pathogens as potential drivers of community assembly, should pathogen load increase with time among these arthropods. At an evolutionary scale, Fukami and colleagues were able to show in 2007 that alternative assembly routes were capable of enhancing and retarding the rate of evolutionary change among community members. Not only does assembly generate local and regional species diversity by creating alternative states, but it also offers nature a variety of scenarios against which evolution proceeds in varied fashion.

Interestingly, recent numerical studies of networks of identical coupled oscillators have shown that pattern in the form of subsets of rogue oscillators (chimera states) are capable of breaking synchrony during system evolution. In systems such as these, additional pattern can emerge if coupled and uncoupled sets respond differently to events like species invasion attempts. It would seem that in light of the extraordinary behaviors recently observed in nonlinear dynamical systems, our null model approach must be recast.

ASSEMBLY OF COMPLEX DYNAMICAL SPACES

Until recently, modeling in ecology focused on the simplest of systems where asymptotic behavior and equilibrium conditions can be precisely evaluated, the assumption being that little of consequence occurs during the transient phase, the period of time between the initiation of dynamics and the systems arrival at equilibrium. This might be true if nature contained one or two species and met the assumptions needed for analytical modeling. Nevertheless, analytical solutions simply do not exist for systems of containing many species, and there is no reason to believe that the behaviors observed here are extensible to systems of higher dimension. In fact, the dynamics seen in many species systems suggest little to no role for a purely analytical approach. At present, numerical simulation or approaches based on graphs and network theory would appear to be our best tools.

The basic process that drives the assembly of any system is coupling or invasion and decoupling or extinction. While it is easy enough to add a row and column to an array, representing an invader colonizing a community, spatial variation in coupling as the invader colonizes is exceedingly difficult to model, yet it is likely an essential feature of assembly.

Clearly, coupling and decoupling have significant consequences for subsequent system development, and herein reside the keys to understanding the genesis of emergent properties and structure. Unfortunately, very little is known about the earliest events in system coupling and decoupling, despite the fact that much of nature spends its existence in this far-from-equilibrium, readily disturbed situation. Immediately following a coupling or decoupling event, a new dynamical system is created that enters into a transient phase, as it begins moving toward its new solution or attractor. However, the dynamical control that previously operated had placed the system somewhere in a particular state space creating unique initial conditions, possibly outside the trajectory field of the new basin. Hence, some period of time exists before the new dynamics fully assume control and pattern exists that reflect those dynamics. During this period there is incomplete coordination or asynchrony among the elements that comprise the new system. Additional coupling to this system during this stage could very well be

successful, but impossible as synchronization proceeds, or vice versa. Colonization success and failure depend not only on species composition and community organization of the community but also on where that community has come from and where it is going.

Consider, for example, a system where a predator population is mediating competitive interactions among its prey, thereby thwarting competitive exclusion and maintaining higher species diversity. Should coupling with another system occur, perhaps through some novel metapopulation or metacommunity interaction or direct invasion, the role of the predator may change. However, the initial distribution and abundance of prey species still reflect previous controls, but now under the control of new dynamical rules. Because nature is spatially extended, expression of this control varies due to pattern previously created, further altering pattern.

Recent numerical work exploring the long-term dynamics of nonlinear systems has produced a tantalizing array of behaviors that appear to have biological analogues. Such studies offer potential explanations for the intricacies of community assembly. For example, attractor ghosts have been discovered that function ostensibly as a switch by delaying system collapse, thus enhancing the possibility of a species invasion and rescuing a system from collapse. Should this occur, a novel state space with new assembly rules emerges, a state potentially derivable only through this route. Similarly, the collective behavior exhibited by coupled oscillators (e.g., metapopulations, metacommunities, population and community patches along some gradient), moving from an unsynchronized, uncorrelated state to a synchronized state, creates a sequence of radically different phenotypes. Here, the reference to phenotypes is based on variation in the response of potential colonists to various stages of synchrony in the system being invaded.

Despite considerable progress, many questions remain unanswered, and needed concepts have yet to be developed. Where is nature going and how might it get there? Can assembly be thought of and cast in terms of self-organization? How do the relative roles of determinism and indeterminism in dynamics drive system assembly? Do dynamical phenomena like basin riddling and ghosting play a role in the developmental trajectories of developing ecological systems? Do assembly processes play a fundamental role in influencing global patterns in species diversity? Finally, could many of the observed regularities seen in nature be a direct function of assembly? For example, the SLOSS (single large or several small) debate in ecology, based on species–area relationships and cast as $S = cA^z$, may have less to do with area than the effect multiple initializations and attendant initial conditions in generating species diversity.

SEE ALSO THE FOLLOWING ARTICLES

Food Webs / Metacommunities / Networks, Ecological / Stability Analysis / Succession

FURTHER READING

Diamond, J. M. 1975. Assembly of species communities. In M. L. Cody and J. M. Diamond, eds. *Ecology and evolution of communities.* Cambridge, MA: Belknap Press.

Dureau de la Malle, A. 1825. Mémoire sur l'alternance ou sur ce problème: la succession alternative dans la reproduction des espèces végétales vivant en société, est-elle une loi générale de la nature. *Annales des sciences naturelles* 15: 353–381.

Elton, C. 1946. Competition and the structure of ecological communities. *Journal of Animal Ecolology* 15: 54–68.

Fukami, T., H. J. E. Beaumont, Xue-Xian Zhang, and P. B. Rainey. 2007. Immigration history controls diversification in experimental adaptive radiation. *Nature* 466: 436–439.

Hutchinson, G. E. 1959. Homage to Santa Rosalia, or why are there so many kinds of animals? *American Naturalist* 93: 145–159.

Petermann, J. S., A. J. F. Fergus, C. Roscher, L. A. Turnbull, A. Weigelt, and B. Schmid. 2010. Biology, chance, or history? The predictable reassembly of temperate grassland communities. *Ecology* 91: 408–421.

Piechnik, D. A., S. P. Lawler, and N. D. Martinez. 2008. Food-web assembly during a classic biogeographic study: species' "trophic breadth" corresponds to colonization order. *Oikos* 117: 665–674.

Simberloff, D. S., and E. O. Wilson 1969. Experimental zoogeography of islands: the colonization of empty islands. *Ecology* 50: 278–296.

Simberloff, D. S., and E. O. Wilson. 1970. Experimental zoogeography of islands: a two-year record of colonization. *Ecology* 51: 934–937.