

CHAPTER TWO

Strategies of Adaptation

The study of adaptation has usually meant the search for the adaptive significance of particular structures or physiological processes. This can sometimes be done directly in laboratory or field studies such as those on protective coloration. A trait can also be demonstrated to be adaptive by showing parallel variation in space or time for several species. In these studies, one assumes a constant environment and relates traits, for instance size to temperature, color to substrate, or hunting behavior to type of prey.

But there are other characteristics of organisms and populations which are not explicable as adaptations to particular environments—the degree of homeostasis in development, the amount of polymorphism, the extent of spatial differentiation of a species, sensitivity to natural selection, degree of inbreeding, and so on. These and other traits may be regarded as adaptations to the pattern of the environment in space and time, to temporal variability, to environmental uncertainty, and to what we will describe below as “grain.” Such adaptations therefore fall into the category of strategies.

When our emphasis shifts to variable environments entirely new problems arise. If large body size is an adaptation to cold by way of the surface/volume relation, what size is optimum in an environment which is sometimes hot and sometimes cold? The commonsense of “folk liberalism” suggests a middle course, an intermediate size between those adapted to the extremes of heat and cold. But as we shall see this is only sometimes the case. In other circumstances a species may adapt completely to one of the extreme environments even at the expense of near lethality in the other. Or it may be polymorphic, containing a mixture of phenotypes (usually genetically determined), some of which are adapted to each environment. Or the animals may arrange to be large in the cold and small in the heat.

In the latter case, the environment may act directly on the developing system to evoke a phenotype which is more adapted to that environment. The response to the environment may be rapid and reversible, such as shivering in the cold, or it may be permanent, such as the size of holometabolous insects which is sensitive to larval temperature and is fixed throughout adult life.

There is always the danger that by the time the adaptation has taken place the environment has changed. This danger is of course greater when the delay is longer, and is especially so when the population adapts from generation to generation by genetic change through response to natural selection. Suppose for example that a species of butterfly has two generations per year, one in summer and one in winter. During the winter, those genotypes which are best adapted to winter conditions survive in greater numbers and leave more offspring to the next generation. Thus the next generation is better adapted to winter conditions than its parents but must develop during the summer. Similarly, in the summer generation those best adapted to summer conditions survive and produce a winter generation which is better adapted to summer. Such a species is always lagging behind the environment, doing the right thing for the previous situation. The result is to reduce fitness even more.

Thus we see that this species cannot adapt to the alternation of seasons by natural selection. However, each generation might adapt developmentally. The cold conditions faced by the larvae of the winter generation could produce cold-adapted adults which survive well in winter, and hot summer conditions may evoke the heat-adapted phenotype in the summer generation.

The same problem also arises in reversible physiological adaptation. Consider an inducible enzyme such as β -galactosidase which is formed to some bacteria in the presence of the sugar galactose. Is it better for the bacteria to contain the enzyme at all times in expectation that it may prove useful sometimes, or to produce the enzyme only in the presence of the substrate, or never to have the enzyme?

The induction of the enzyme takes time, roughly on the order of half an hour. Thus the cell which always has the enzyme gains about 30 minutes of growth whenever galactose appears as against those cells which must first synthesize it. However, it must tie up energy and protein in the form of an inactive enzyme when galactose is absent.

If the galactose is present or absent in bursts of 10 hours, the inducible cells lose 5% of the opportunity for growth, but pay only half the cost. Unless the cost is negligible it would seem that induction is the better strategy. But if galactose appears and is removed in periods of 30 minutes, the inducible cells still pay half the cost but now derive no benefit. In this case the permanent enzyme is the better strategy. Finally, if short spurts of galactose occur rarely enough it may be better never to utilize it.

This suggests that it would be advantageous to speed the rate of response. If the enzyme could be fully synthesized in one minute, then half-hour fluctuations in the presence or absence of galactose could be met by induction. The environment at the time of induction would now be a good prediction of the environment after the response is completed.

Suppose, however, that galactose is introduced into the medium either in spurts of one minute or for ten minutes at a time. If the enzyme is induced after one minute in the presence of galactose, and if the short spurts are more frequent than the longer ones, the cost of synthesis may exceed the benefit derived from using galactose when it remains for ten minutes. Here perhaps a two-minute lag might be optimal. Now the cells lose 20% of the growth time in the long spurts but cut the cost in half by not responding to the short spurts. The optimal strategy must be able to distinguish between "signal" and "noise," doing so in this case by their different durations.

In the above cases the environmental factor to which the response is an adaptation is also the signal that evokes the response. But this need not be the case. Many insects go into a dormant diapause during the winter as an adaptation to the scarcity of food or to the cold. This diapause may be obligatory, in which case the insect is univoltine and has

only one generation per year, or facultative, in which case the insect is multivoltine and only the last generation of the summer enters diapause. The determination of whether an animal will go into diapause or not often occurs in the egg. Although it may be an adaptation to the cold, the temperature over a few days is not a very good prediction of the temperature next month. The day length is generally the best single predictor, providing the best information as to the season. This brings us to the problem of environmental information in general.

The organism is conceived of as receiving this information through several receptor systems or through its whole body, processing this information into a prediction of future environments, and responding to this prediction with the formation of that phenotype which is the best strategy for the predicted range of environments. There is no necessary relation between the physical form of the information and the environmental factor to which the organism is responding. What is important is the statistical relation between them. Thus for most multivoltine insects a long day indicates summer and hence evokes direct development while a short day indicates autumn and evokes diapause. But in the silkworm moth, *Bombyx mori*, there are only two generations per year. The first is formed in early spring when the days are still short. It develops immediately in the short day and produces the second generation in summer when the days are longest. But here the long day predicts winter and brings on diapause.

The hypothesis that there is no necessary relation between the physical form of the signal and the response evoked is based on the assumption that there is a complex network of causal relations between the point of contact with the environment and the responding system. Different loops of this network can have opposite effects on the reacting system, some enhancing and some suppressing the original reaction. Therefore selection of an optimum response comes about by increasing some pathways and blocking others. The hypothesis is a useful one and suggests interesting experiments, but it will not hold universally. Where the

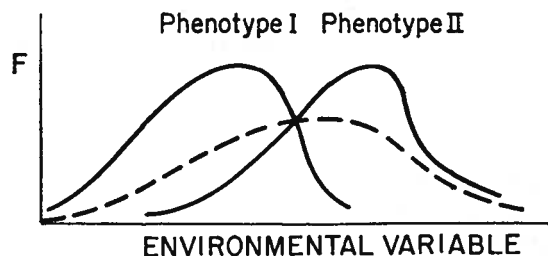


FIGURE 2.1a. Fitness as a function of the environment for two phenotypes. The dotted curve is the fitness of a mixture of the two phenotypes. The peak occurs at the optimal environment for each phenotype.

pathway between the environment and the responding system is short, its evolutionary flexibility will be much reduced. This seems to be the case for temperature. High temperature always accelerates development in the fully viable range and results in smaller body size in invertebrates, even where large size would seem advantageous. Some of the consequences of this will be discussed in Chapter 4.

We will now introduce the method of fitness sets for the analysis of adaptive strategy. In Figure 2.1a we show the relation between a component of fitness and the environment for an organism which may be in one of two physiological states. The location of the peak gives the optimal environment for that phenotype, the height of the peak is the measure of the best performance in the optimal environment, and the breadth of the curve is a measure of the tolerance for non-optimal environments. This tolerance is a measure of homeostasis. If there were no restrictions on the curve in Figure 2.1a, the optimal curve would obviously be infinitely high and infinitely broad. In fact there are restrictions. The height at the peak is undoubtedly limited by the physico-chemical structures involved. But in addition we suggest that the breadth cannot be increased without lowering the height. Suppose that the phenotypes I and II in the figure refer to two enzymes with different pH or temperature optima but with curves of the same shape. If the same total amount of enzyme is divided equally between the two forms

the combined system has a fitness curve shown by the broken line in Figure 2.1a. If the enzymes are mixed in any proportions the shape of the curve may alter but the total area under the curve remains constant. If the curves refer to different genotypes in the same population the same argument holds.

The researches of George Sacher (1966) allow us to extend this principle of allocation further. It was first observed that despite differences in the total life spans of different animals, the caloric life spans measured in energy expenditure are remarkably uniform. The major discrepancies from constant expenditure were associated with the homeostatic system—the bigger the brain the lower the rate of aging per calorie. Thus Sacher defines organizational entropy as a measure of the energy cost per unit of development (carrying the organism from one stage to another). He amassed a great deal of data on his $S_{org}(T)$ as a function of temperature. It has a minimum value at some optimal temperature and increases with the departure of the temperature from the optimum. This “entropy” curve is broader and flatter for insect eggs and pupae than for larvae, which being mobile can seek out preferred environments. Sacher’s work therefore suggests that the cybernetic system which reduces the organizational entropy is itself costly. It can be extended to cover a wider range of environments, but only by reducing the efficiency at the optimum.

Thus we assert the principle of allocation: the fitness curve $W(s)$ for an environmental parameter s may vary in shape but is subject to the constraint

$$\int F\{W(s)\} ds = C. \quad (2.1)$$

We do not know the functional form of F in general, but where fitness is altered by mixing components such as enzymes, $F(W) = W$.

For the purposes of this study we can specify that for any phenotype y and environment s , $W(s - y)$ is a non-negative function with a maximum at $s = y$ and decreasing symmetrically toward zero as $|s - y|$ increases. The dual of the curve in Figure 2.1a for a fitness component over environ-

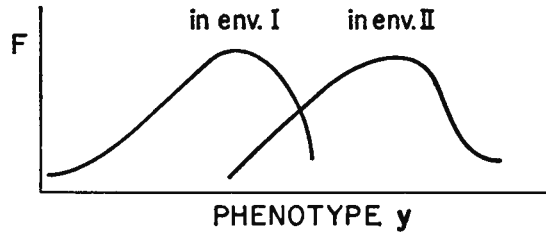


FIGURE 2.1b. Fitness as a function of phenotype in two environments. The peak occurs at the optimal phenotype for each environment.

ments is shown in Figure 2.1b, where for a fixed s we plot $W(s - y)$ over a range of phenotypes y . For any two environments the curves overlap. If they are close enough so that their inflection points overlap, the average of the two curves (fitness in an environment which is half S_1 , half S_2) will have a single peak in the middle. If S_1 and S_2 are farther apart the average curve has a minimum at the midpoint and two peaks near S_1 and S_2 .

The fitness set representation presents the curves of Figure 2.1b in a different way. The two axes in the graph are now fitness components W_1 and W_2 in environments S_1 and S_2 . The phenotype whose peak is at S_1 in Figure 2.1b gives the point farthest to the right in Figure 2.2. The phenotype corresponding to S_2 gives the uppermost point. All pheno-

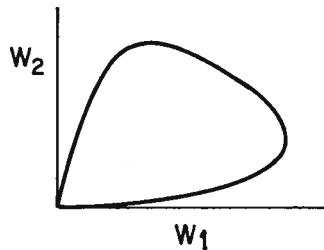


FIGURE 2.2. The fitness set. The coordinates of each point on the fitness set are the ordinates in Figure 2.1b, corresponding to the two environments for each phenotype.

types which have curves differing only in the location of their peaks will lie on the boundary of the fitness set shown in Figure 2.2. Those for which the area under the curve is less will lie inside the fitness set.

It can readily be seen that if S_1 and S_2 are sufficiently close (so that their inflection points overlap) the fitness set will be convex along the upper-right-hand boundary, while if S_1 and S_2 are farther apart the upper-right-hand boundary will be partly concave. This difference between the concavity and convexity of the fitness set will have important biological consequences. Therefore we define twice the distance from the peak to the inflection point as the tolerance of the phenotype and assert that the fitness set is convex or concave depending on whether the environmental range $|S_1 - S_2|$ is less than or greater than the tolerance of a single phenotype.

The importance of this distinction is that the fitness of a mixture of phenotypes in a population or of physiological states in an individual is represented by a point on the straight line joining their points on the fitness set. In a convex fitness set such mixed strategies will lie inside the set and therefore each one will be inferior to some single phenotype which lies above and to the right of it. But on a concave fitness set certain mixtures will lie outside (up and to the right) of the fitness set for single phenotypes so that mixed strategies may be optimal.

The fitness set alone does not define an optimum strategy. Over-all fitness in a heterogeneous environment depends on the fitnesses in the separate environments, but in a way which is determined by the pattern of environments. We therefore define the Adaptive Function $A(W_1, W_2)$, which measures fitness in the heterogeneous environment, to be a monotonic increasing function of its arguments. If the environment is sometimes S_1 and sometimes S_2 , the individual must survive in both in order to survive. Let the probability of dying (or the loss in growth rate) in the interval Δt be $m(t) \Delta t$ where $m(t)$ takes on two values according to which environment S_1 or S_2 is currently present. Then the probability of survival up to time t is $P(t)$, and it satisfies

the relation

$$P(t + \Delta t) = P(t)[1 - m(t) \Delta t]$$

(assumes $m(t)$ unchanging for at least Δt). (2.2)

Where Δt is a whole generation,

$$P(t) = \Pi(1 - m(t) \Delta t) \quad (2.3)$$

which is the product $W_1^p W_2^{1-p}$ for environments S_1 and S_2 occurring in the proportions $p:1 - p$. But if Δt is very small the terms in $(\Delta t)^2$ and higher powers vanish, and

$$P(t) = P_0 e^{-\int m(t) dt}. \quad (2.4)$$

This will be maximized when the integral $\int m(t) dt$ is smallest, which occurs when the linear average of the fitnesses $pW_1 + (1 - p)W_2$ is greatest. In the former case the environment is described as coarse-grained and presents itself to the individual as alternatives. Thus a coarse-grained environment is uncertain for the individual even if the proportions of the alternative environments remain fixed. In the latter case the environment is experienced as a succession of possibly different conditions. The differences present themselves to the organism as an average which is the same for all members of the population. There is no uncertainty. Of course there may be intermediate-grained environments, with the Adaptive Function intermediate between the linear and the multiplicative (or logarithmic).

The notion of grain comes of course from the size of patches of environment. If the patch is large enough so that the individual spends his whole life in a single patch, the grain is coarse, while if the patches are small enough so that the individual wanders among many patches the environment is fine-grained. But the concept can be made more general. For example, since most animals eat many times, food differences are fine-grained, whereas alternative hosts, when several are available for a parasite, are coarse-grained differences.

Thus for the situations we have described the Adaptive Function $A(W_1, W_2)$ may vary from the hyperbola-like $W_1^p W_2^{1-p}$ to the linear $pW_1 + (1 - p)W_2$. In any case the

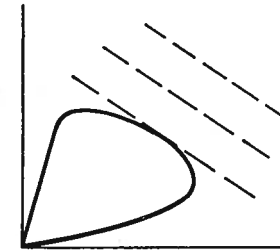


FIGURE 2.3a. Fine-grained environment on a convex fitness set.

optimal strategy is represented by the point on the fitness set (single phenotype or mixed strategy) which touches the curve $A(W_1, W_2) = K$ at the greatest K . In Figures 2.3a, 2.3b, and 2.3c we show several optimum strategies. In formal terms the results are the following:

1. On a convex fitness set (environmental range smaller than the tolerance) the optimum strategy is a single phenotype which is adapted to some intermediate value of S between S_1 and S_2 , and does moderately well in both these environments.
2. On a concave fitness set (environmental range exceeds the tolerance) a fine-grained environment results in an optimum strategy of a single phenotype which is specialized to either environment S_1 or S_2 depending on p , the frequency of environment I. Therefore it does optimally in one environment and poorly in the other.
3. On a concave fitness set with a coarse-grained environment the optimum is a mixed strategy in which the two

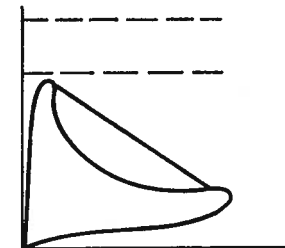


FIGURE 2.3b. Fine-grained environment on a concave fitness set.

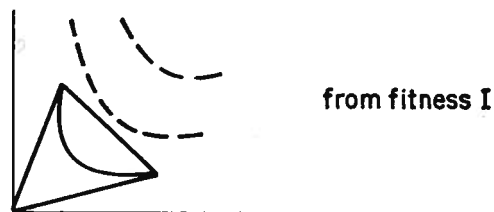


FIGURE 2.3c. Coarse-grained environment on a concave fitness set.

specialized phenotypes occur in proportions that depend on p .

When there are more than two environments new axes have to be added, and in a continuous environment we need infinitely many dimensions for the fitness set. However, as S_1 becomes arbitrarily close to S_2 any phenotype which does well in one will do well in the other, and the fitness set becomes a straight-line segment at 45° from the origin. Thus we only have to consider as distinct those environments different enough to change the ranking of fitnesses of the available phenotypes.

An alternative approach considers the environmentally determined optimum phenotype S to have a probability distribution $P(S)$, and attempts to maximize $\int A[W(S)]P(S) dS$ subject to the restriction $\int W(S) dS = C$. When $A(W)$ is linear, the optimum allocation is specialization to the most common S , while for a coarse-grained environment $A(W)$ is $W_1^p W_2^{1-p}$ or equivalently $p \log W_1 + (1-p) \log W_2$, and the optimum is $W(S) = CP(S)$. Thus the fitness is assigned to each environment in proportion to its frequency. Qualitatively the results are the same—coarse-grained environments are uncertain and give rise to less specialized broad-niched populations, with the niche breadth increased by the uncertainty of the environment. The niche breadth interpretation is pursued further in Chapter 3.

On the basis of this formal theory we can now interpret biological situations by defining "phenotype" in various ways.

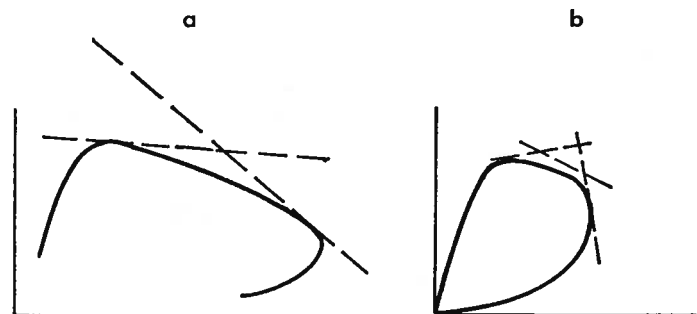


FIGURE 2.4a. Optimum phenotype as a function of the probability of a particular environment p_e . A small change in p_e alters the slope of the Adaptive Function, moving the optimum phenotype a long distance on a flat fitness set.

FIGURE 2.4b. Optimum phenotype as a function of the probability of a particular environment p_e . Here a small change in the position of a point on the fitness set corresponds to a large change of slope.

A developing organism may respond to a variable environment in several ways:

1. It may always produce the same phenotype, regardless of the environment. This corresponds to Schmalhausen's (1949) autonomous development and Waddington's (1957) canalization. This would be the optimum strategy in the absence of environmental information—when fluctuations are quite rapid compared to the generation or when there is no signal that is predictive of future environments and which the organism can process effectively. This stability of a given phenotype may of course depend on some internal strategies of other kinds which minimize the effects of environmental perturbations.

2. It may produce a phenotype which varies as a continuous function of the environmental signal (Schmalhausen's dependent development). In Figure 2.4 we show how the optimum phenotype varies with the probability of environment I, p_e . The subscript indicates that p_e is a conditional probability, the probability of environment I given the environmental signal e . But the phenotype changes at a rate which depends on the shape of the fitness set. If the

fitness set is more or less flat, a small change in p_e will move the point of tangency a great distance along the fitness set, which means that a large change of phenotype is involved. But for a very convex fitness set as shown in Figure 2.4b almost all slopes p_e can be found along a short segment of the fitness set, so that in the range near $p_e = 0.5$ there is only slight change in the optimum phenotype with the environment. The sensitivity of the optimal phenotype to the environment also depends on the grain. In a fine-grained environment the slope of the Adaptive Function is

$$\partial W_1 / \partial W_2 = -(1 - p) / p \quad (2.5)$$

so that the derivative of the slope with respect to p_e is

$$\partial(\partial W_1 / \partial W_2) / \partial p = 1 / p^2. \quad (2.6)$$

For a coarse-grained environment the corresponding derivative is

$$\partial(\partial W_1 / \partial W_2) / \partial p = (1 / p^2) W_1 / W_2. \quad (2.7)$$

Since p is the proportion of environment I, when p is large W_1 / W_2 is also large. As p decreases, $1 / p^2$ increases but W_1 / W_2 decreases, so that a given change in p produces a smaller change in the optimum phenotype.

3. It may produce one phenotype for environments below a given threshold and a second phenotype above that threshold (Schmalhausen's autonomous regulative development, or developmental switch). This is the optimum strategy for a fine-grained environment when the range of environment exceeds the tolerance of the individual phenotype.

4. It may produce a mixture of phenotypes, the probability of each depending on the environment. This type of development, a stochastic switch, would be optimal for coarse-grained environments when the range exceeds the individual tolerance.

Patterns 1 and 2 are the most familiar ones. While they are understandable in a general way, the quantitative interpretation of rates and degrees of acclimatization is not obvi-

ous. We know for example that temperature acclimation requires somewhat less than one day for amphibians (Brattstrom and Lawrence, 1962) and for *Drosophila* (our work), but substantially longer for fish. Also, different species acclimate to different degrees. Among the Puerto Rican *Drosophila*, the narrow-niched species were all poor at acclimation to heat. But the broad-niched species included two which acclimate well, one which had a very high heat resistance without acclimation, and one which hardly acclimated at all.

The switch patterns are also widespread. They include the determination of sexual vs. asexual eggs in the Rotifera, dormancy in many seed plants, diapause in insects and mites, the production of winged vs. wingless aphids, formation of free-living vs. infective larvae in some nematodes including Strongyloides, aerial vs. aquatic leaves in some plants, etc.

E. O. Wilson (unpublished) used the dual form of the fitness set to analyze caste polymorphism in ants. We transpose his results here into our fitness set notation. Wilson considers that an ant colony faces many crises of different kinds that can reduce the production of queens, which is the ultimate measure of fitness. Different phenotypes can cope differentially with these situations. For any two situations the fitness set may be represented as in Figure 2.5. When the fitness set is convex, the optimum is a single all-purpose caste. But for a concave fitness set (which means, for very different situations) polymorphism may be optimal. Wilson postulates that selection pressure will be greatest on those traits involved in meeting the most serious or most frequent contingencies, so that finally for all contingencies i with frequency p_i and fitness contribution W_i , $p_i W_i = \text{constant}$. Thus his Adaptive Function maximizes the total fitness $\sum p_i W_i$ subject to the restriction $p_i W_i = p_j W_j$. Thus the optimum population is composed of that mixture of casts whose fitness point lies on the line $p_1 W_1 = p_2 W_2$ where it crosses the boundary of the fitness set. Unspecialized castes are roughly equally fit for both contingencies and therefore are represented by points near the 45° line, while specialized

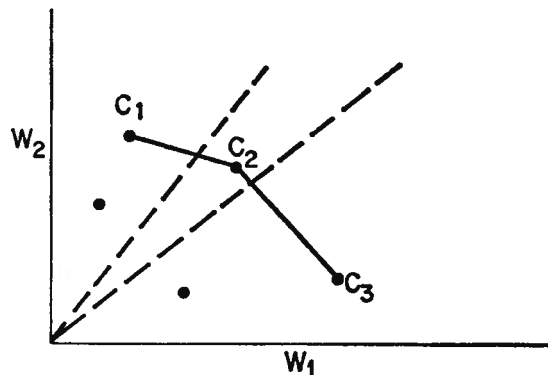


FIGURE 2.5a. A change in the relative weights assigned to different contingencies may result in the replacement of one caste by another.

castes lie closer to the two axes. From these considerations the following can be proved:

1. The number of castes cannot exceed the number of kinds of contingencies to be met. This follows from the fact that the line $p_1W_1 = p_2W_2$ always passes between a pair of

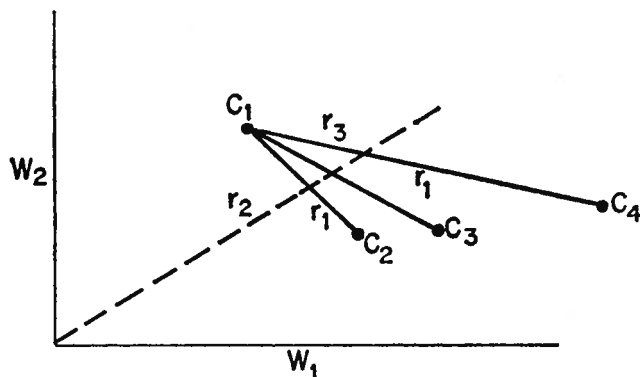


FIGURE 2.5b. The optimal population consists of two castes in the proportions of the line segments divided by the Adaptive Function (dotted line). Thus castes C_1 and C_2 are more or less equally common, but a population of C_1 and the more specialized C_3 will be mostly C_1 .

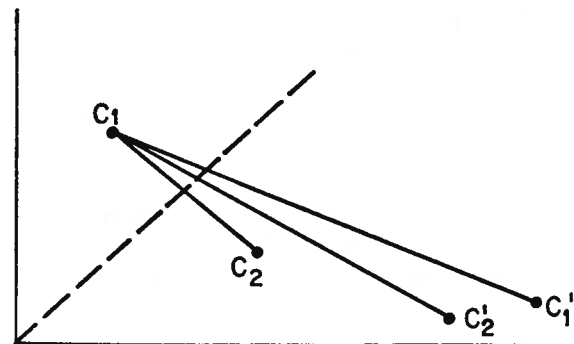


FIGURE 2.5c. Selection favors increased specialization. As C_2 increases in specialization its fitness point moves to the right. The mixed populations form a new fitness set which is superior to the previous one.

fitness points. (We exclude the infinitely unlikely event that three fitness points are colinear.)

2. A change in p will produce greater changes in caste proportions if these are unspecialized (near the 45° line). A pair of specialized castes is a more stable arrangement, less likely to be broken down during evolution, than a pair of unspecialized castes. But as Figure 2.5a shows, in a polymorphism of one unspecialized and one specialized caste the unspecialized caste is retained as specialized castes replace each other.

3. The more efficient a caste at meeting the contingency for which it is specialized, the lower its abundance. This is seen in Figure 2.5b.

4. Increased specialization of a caste, improving its fitness at one task, usually increases colony fitness even when it reduces the effectiveness of that caste for other tasks. This is shown in Figure 2.5c.

Strategic analysis has been applied by Martin Cody (1966) to the problem of clutch size in birds. He uses a three-dimensional fitness set whose axes are clutch size, predator avoidance, and competitive ability. He assumes a convex fitness set so that it is not necessary to specify the Adaptive Function at all except that it increases along each axis. Since clutch size would be an aspect of r -selection (selection for the rate of population increase when it is below satura-

tion), it will have greater weight in environments which are unstable from generation to generation, and populations may often be below saturation. This leads to the prediction that clutch size will increase not only from the tropics to the temperate zone (which is already known) but also from the coast inland and from sea level to higher elevations. Further, tropical regions with great variation in resources such as are found for sea birds off the coast of Peru will also show large clutch size compared to close relatives elsewhere.

Toward regions of greater environmental stability the rate of predation is usually greater, so that the maximization of the Adaptive Function would result in decreasing clutch size but an increase in the anti-predation expenditure. However, for birds which are well protected by nesting in holes, and for many island dwelling birds, predation is much reduced, and the tropics-temperate gradient in clutch size will therefore be less steep.

Cody considered a single fitness set for all birds. A more detailed study of the biology of different families would show predictable differences in the convexity of the fitness set, which would affect the steepness of the gradient. In fact, the case of the hole-nesting species could be looked at in this way. The choice of a nest site in a hole does not reduce the ability to feed. Therefore, when predator avoidance is based on hole nesting instead of on making fewer visits to the nest, the fitness set is more convex on the predator-avoidance-clutch-size plane. Therefore a given change in the slope of the Adaptive Function would result in a smaller change in clutch size. Instead, Cody allowed hole nesting to reduce the weight of predator avoidance in the Adaptive Function. Both methods are equally valid. The division of a biological problem into the fitness set and the Adaptive Function is to some extent arbitrary. We have found that it is usually most convenient to divide the difficulties of the problem more or less equally between the two.

The problem of delayed reproduction is a widespread one, including delayed maturity, seed dormancy, resting spores in some invertebrates, and diapause. We can treat the diapause problem as representative of these questions.

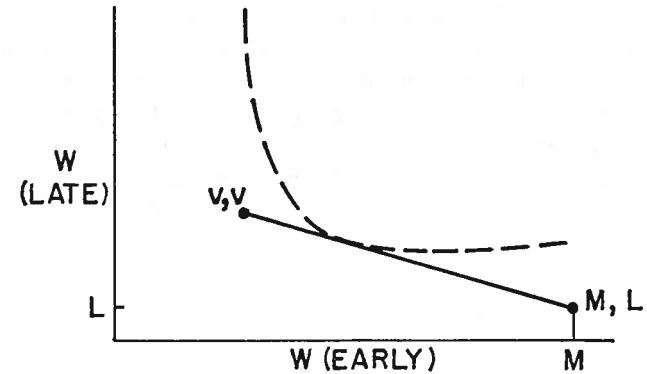


FIGURE 2.6. A fitness set analysis of diapause strategy. The point v, v represents diapausing forms, while M, L represents the non-diapausing phenotype.

Let the axes of the fitness set be fitness in early summer (when there is time for another generation) and in late summer (when there is no longer time). The phenotypes are duration of diapause, from zero to several months. The intermediate-length diapause reduces fitness in both environments—reproduction is lower due to the delay but the breaking of diapause in midwinter is also lethal. Therefore the fitness set is concave, and we only have to consider the two phenotypes complete versus no diapause.

The dormant eggs have a small probability of death, which for convenience we assume to be about the same in early and late summer. The non-dormant eggs are virtually inviable in late summer but will leave many offspring if it is early summer. Thus the fitness set is highly asymmetric. This is shown in Figure 2.6.

The alternative environments are “coarse-grained” so that the Adaptive Function is $p \log W_1 + (1 - p) \log W_2$. It has a slope of $-(1 - p)W_1/pW_2$, where p is the probability of environment I. The relevant part of the fitness set is the line joining the points for diapause and non-diapause. If v is the viability of dormant eggs, M is the viability of non-dormant eggs in early summer, and L is the viability of non-dormant eggs in late summer, the fitness set has a slope of $-(M - v)/(v - L)$. The optimum strategy may

be all eggs developing directly, all in diapause, or some fraction of the eggs in each class. Let the proportion of eggs developing directly without diapause be q . Then the rate of increase if good conditions prevail is $\log \{v + q(M - v)\}$ and if winter comes within a generation, fitness is $\log \{v - q(v - L)\}$. Thus the average fitness is $p \log \{v + q(M - v)\} + (1 - p) \log \{v - q(v - L)\}$.

Hence:

$$\hat{q} = v \left(\frac{p}{v - L} - \frac{(1 - p)}{M - v} \right)$$

provided that this lies between 0 and 1. Hence:

$$\hat{q} = \begin{cases} 0 & p \leq \frac{v - L}{M - L} \\ v \left(\frac{p}{v - L} - \frac{(1 - p)}{M - v} \right) & \text{for } \frac{v - L}{M - L} \leq p \leq \frac{M}{v} \left(\frac{v - L}{M - L} \right) \\ 1 & p > \frac{M}{v} \left(\frac{v - L}{M - L} \right) \end{cases} \quad (2.8)$$

Therefore for $M = 1$, $L = .1$, $v = .9$, the optimum q is 0 when p is less than $.8/.9$ or about $.89$ and rises to 1 when $p = .8/.9 \times 1/.9$ or about $.97$. For a large value of M the transition from no diapause to complete diapause is less abrupt. For $M = 10$ and the other terms unchanged, $\hat{q} = 0$ when p is below $.8/9.9 = .08$, and rises to 1 at $p = .89$. Thus the same kind of adaptive system can result in the relatively abrupt switch from 0 to 100% diapause, or can produce a slower gradual transition as in the production of winged vs. wingless aphids.

Dan Cohen (1967b) has approached the problem of seed dormancy analytically and numerically. Among his conclusions is the observation that an optimizing system would not generally utilize all the available environmental information. Signals that would evoke the same response can be lumped, and a new source of information will be used only if it would lead to different responses which increase fitness

sufficiently to offset the cost of reception. In another study the same author (1967) analyzes the sufficient signal for bird migration.

Monte Lloyd (1966) considered a very unusual kind of delayed reproduction—the 13-year and 17-year generations in the periodic cicadas. While the origin and maintenance of this pattern is difficult to account for, the adaptive significance seems to be the escape from control by predators. A predator whose life cycle is less than the generation length of the cicada would be out of phase with the cicada years most of the time, since 13 and 17 are prime numbers.

Strategic analysis can also be applied to behavior. Here we consider only one aspect of behavior: active habitat selection. Consider a species in which the individual has a limited time in which to find a suitable microhabitat. Let there be two kinds of usable sites, a preferred one with viability M and a less satisfactory site where fitness is L . The density of habitat type I is p , so that the probability of finding one such site in time t is $1 - e^{-\lambda p t}$, where λ is a measure of searching ability or mobility and t is the available time. The density of habitat II is $q = 1 - p$. However, if this habitat is found it will only be used with a probability θ . The problem is to optimize θ .

Total fitness is the fitness if a site is found times the probability of finding a site. This is

$$W = \left[\frac{pM}{p + \theta q} + \frac{\theta q L}{p + \theta q} \right] [1 - e^{-\lambda(p + \theta q)t}]. \quad (2.9)$$

The first bracket is equal to $M - (M - L)\theta q / (p + \theta q)$ and therefore decreases with θ . The second term increases with θ . Clearly, if $M - L$ is negligible the optimum $\theta = 1$, while if λt is large enough the second term will be near 1 regardless of θ , and the optimum θ would be zero.

In a heterogeneous population, $M - L$ depends on the gene frequency. For each gene frequency there is a value of $M - L$, and this determines the optimum θ . But θ , by determining the effective proportions of the two environments which the population actually faces, can determine

gene frequency. Thus there is a coevolution of specialization and habitat preference, which will be examined in Chapter 3.

We see that for a fixed $M - L$, the degree of habitat specialization is determined by $1 - e^{-\lambda}$, the probability of finding a usable site. This is a measure of the effective productivity of the environment for the species in question. Low productivity is in effect uncertainty and, like other kinds of uncertainty, results in a broadening of the niche (a reduction of specialization).

MacArthur and Pianka (1966) have applied the same kind of argument to hunting behavior. They distinguish between searching and pursuing. During the search the hunter is not committed to any particular kind of prey (except indirectly by selecting the habitat where it searches). In contrast, pursuit implies the chasing of a particular food item and therefore precludes hunting for anything else. MacArthur and Pianka reason that if the search time is long compared to the pursuit time, very little search opportunity is lost during pursuit. It is therefore advantageous to pursue any capturable item. But if pursuit is long compared to search the energy invested in pursuit might be better invested in searching for the preferred items.

Hence pursuers may specialize more readily in their choice of food than searchers can. The relative importance of searching and pursuing is partly a property of the biology of the animal and partly a property of the abundance of food. Once again high productivity leads to low uncertainty and specialization.

We now pass to a consideration of population strategies based on genetic make up and the response to natural selection. Here, in addition to the principle of allocation, there are constraints imposed by the genetic system. Some of these will be discussed in Chapter 5. Here we note only the following:

1. If the fitness set is convex, the optimum strategy is monomorphic. But the optimum phenotype may vary with a slowly changing environment. The population can either remain monomorphic at that phenotype which is the aver-

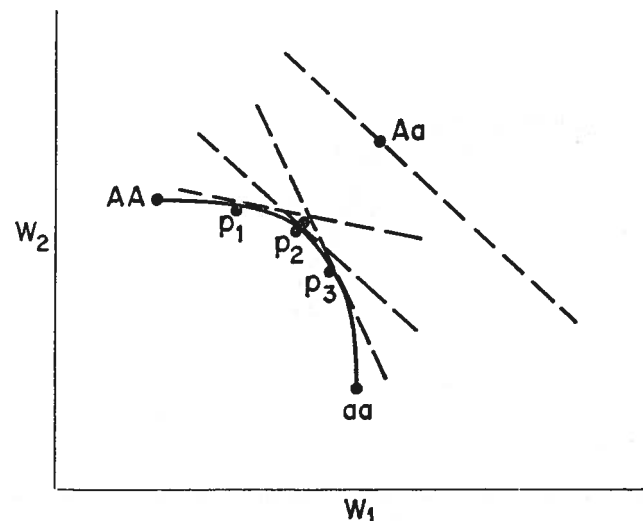


FIGURE 2.7a. Fitness set with genetic constraints. Average heterosis permits polymorphism in a fine-grained environment provided both environments are abundant enough (the Adaptive Function not too horizontal or vertical).

age optimum, or change in response to selection. But the response to selection requires some genetic variance within the population, thus preventing complete monomorphism.

2. A stable polymorphism of specified optimal types may segregate non-optimal types as well.

Both of these departures from optimality are part of the cost of maintaining a strategy.

Consider first a habitat with two environments. Let a population show variation for a phenotype controlled by a single locus with two alleles. The fitnesses of the three possible genotypes can be shown on the fitness set in Figure 2.7. Any panmictic population can be represented by a point on the curve (AA,aa). This curve bisects the median from the heterozygous vertex Aa to the base AA,aa of the triangle in the figure. Inbreeding has the effect of flattening the curve toward the base AA,aa.

In a fine-grained environment the Adaptive Function is $pW_1 + (1 - p)W_2$. Therefore (Figures 2.7a,b) polymorphism will only be optimal when there is average hetero-

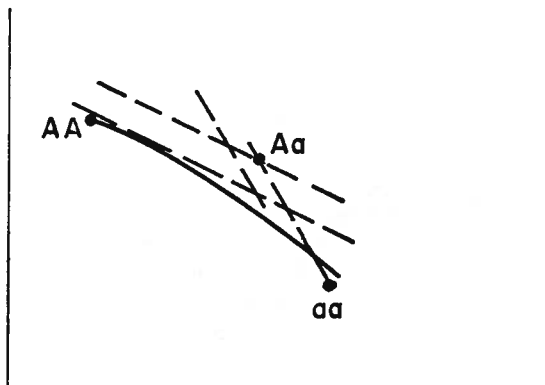


FIGURE 2.7b. With a flatter fitness set the range of polymorphism is restricted.

sis (that is, Aa is above and to the right of the line joining the homozygous points). Further, if p is too close to 0 or 1 the population will be homozygous. Unlike the previous fitness-set arguments, an environment which is too rare will have no effect on the optimum.

Where there is polymorphism in this model, it is not an optimum mixed strategy in the sense of the previous discussion. The optimum would still be monomorphic, of the heterozygous phenotype. But a population of all heterozygotes is not normally possible, so that the polymorphism is imposed by segregation and in fact reduces fitness. In Figure 2.7c we see the segregation load.

With fine-grained environment and a concave fitness set, one or the other allele is fixed. There is therefore no continuing response to selection for different environmental frequencies, whereas on a convex fitness set each frequency of environments determines a gene frequency. The sensitivity of gene frequency to the environment depends on the shape of the fitness set. In Figure 2.7a we see a very convex set. As we argued previously, a change in the slope of the Adaptive Function moves the equilibrium point (which is also the optimum strategy subject to the constraints of the genetic situation) along the fitness set to a point with the same slope. For a highly convex set, this is accomplished

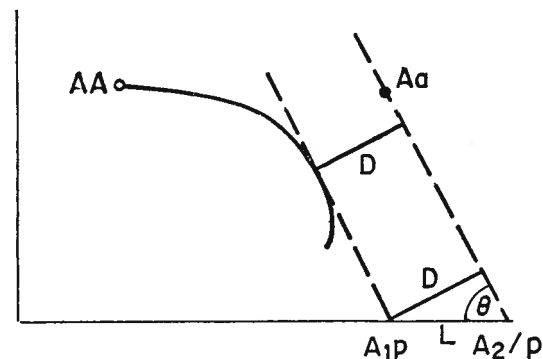


FIGURE 2.7c. The fitness loss due to segregation pL . $D/L = \cos \theta$, and $\tan \theta = -p/(1-p)$.

along a relatively short arc. Thus even wide fluctuations in the environment will produce only small changes in gene frequency. This polymorphism will be highly stable. Such a situation is of course advantageous when there is little correlation between the environments of successive generations so that the optimum gene frequency is hardly altered. But note that this stability is achieved at the cost of a large segregation load.

In Figure 2.7b the fitness set is much flatter. Here the loss of fitness due to segregation is less, but the gene frequency varies more as the environment changes. The fitness set can be flattened by reducing the fitness of the heterozygote or by lengthening the line AA,aa . This is equivalent to increasing the difference between homozygotes.

In a coarse-grained environment the situation is different. It has been discussed in Levins and MacArthur (1966) and will be examined in subsequent chapters.

A more precise description of the behavior of a genetic system under selection can be gotten from the model in which there is an optimum phenotype S which is an environmentally controlled random variable. For each phenotype y , fitness declines with the square deviation from optimum so that

$$W = 1 - (S - y)^2/H. \quad (2.10)$$

TABLE 2.1. Genetic model for selection in a varying environment

Genotype	Frequency	Phenotype	Fitness
AA	x^2	a	$1 - (s - a)^2/H$
AA'	$2x(1 - x)$	0	$1 - s^2/H$
A'A'	$(1 - x)^2$	$-a$	$1 - (s + a)^2/H$

Here H is a measure of homeostasis. The model for a single locus with two alleles and no dominance is shown in Table 2.1. The parameter a is the average effect of the allele A on the phenotype. The fitness of the population at any one time is

$$\bar{W} = 1 - (S - M)^2/H - VAPHE/H, \quad (2.11)$$

where M is the mean phenotype of the population at that time and $VAPHE$ is the genetic variance at that time. The average fitness over time is then seen to be

$$E(\bar{W}) = 1 - (1/H)(\sigma_S^2 + \sigma_M^2 + E(VAPHE) - 2 \text{cov}(S, M)). \quad (2.12)$$

Thus fitness is reduced by environmental variance, by genetic change (σ_M^2), and by the average variance in the population. The first component is imposed from without, but the other two are part of the cost of strategy. This fitness loss may be offset by the covariance term provided the correlation between the environments of the successive generations is high enough (for this model, about 0.8).

It was shown (Levins, 1965) that the amount of genetic variance needed to carry out the optimum tracking of the environment is quite small, at least an order of magnitude less than the variance of the environment. Further, the optimum strategy is not very efficient. It only restored up to one-third of the fitness lost by environmental fluctuation. Of course other models might give more efficient tracking.

The approach outlined here places environmental heterogeneity at the center of our analysis. A number of environmental properties are identified as sufficient parameters: the environmental range, the uncertainty, the grain, tem-

poral variance, autocorrelation between environments of successive periods, and cross correlations between environmental factors which allow signals such as photoperiod to act as predictors of physically unrelated factors.

We also specified several properties of the organism which are taken as given: the fitness set, which gives the relation between fitnesses in different environments for each potential phenotype; the information-receiving capacity of the system; and in some cases the genetics. These organismic properties interact with the environmental ones to give the optimal strategies. These are monomorphic specialization, monomorphic generalization (intermediate in both environments), fixed polymorphism (mixed strategy), and environmental tracking.

The optimum strategy may be found at each of several levels. Once established, it determines the fitness set for the next higher level. Thus a mixed strategy in the sense of a mixture of isoenzymes results in a broader tolerance of the synthetic process involved than can be achieved by one enzyme type alone. This would make the fitness set for cellular types more convex. Thus adaptations occur in an interlocking hierarchy at various levels within the individual physiology, through behavior, and at the level of population structure and dynamics. But at each level the same general principles emerge:

1. Environmental heterogeneity reduces fitness below what it would be in a uniform environment.
2. An optimum strategy can be found subject to the restrictions imposed by the biological situation. This optimum strategy reduces but cannot eliminate the fitness loss due to environmental heterogeneity.
3. The optimum strategy itself imposes a cost on the system. Therefore it is not enough that an adaptive mechanism works—it must increase fitness in terms of the original phenotype-environment system by more than some threshold amount (the cost) before it will be advantageous.
4. At all levels, increased environmental uncertainty results in optimum strategies which spread fitness over a

broader range of environments. If the environmental range exceeds the capacity of a single adaptive system, a mixed strategy will be optimum at that level.

There are a number of major gaps in the theory. For instance, we have been assuming all along a single environmental component and a single corresponding phenotype. But of course most adaptive characters are relevant to survival in several ways, and each component of fitness is affected by many aspects of phenotype. Formally we could assert that environment and phenotype are both vectors rather than scalars, but that would obscure a number of possibilities that we will not explore further now but merely list:

1. A mixed strategy may involve different parts of the phenotype. For example, a plant which is surviving in rain forest and also in a drier forest may have the large rounded leaves typical of the first environment but drop them seasonally as befits plants of the deciduous forest. Or it may have the opposite, with evergreen habit but a more xeric leaf form. There are many possible combinations which would result in great morphological variety even in the same kind of strategy.

2. This creates new possibilities for shifting a trait from one adaptive context to another. For instance, in the anoline lizards pigmentation is relevant to temperature relations and to camouflage. If a dark pigmentation allows an animal to warm up quickly in a cool habitat, and also matches the color there, the same trait is doing double duty and there is no problem. But if the habitat is cool and the substrate is light then what is helpful with respect to temperature is harmful for predator avoidance. Unpublished observations by Harold Heatwole suggest that the conflict is partly resolved by behavior. He found that the better-camouflaged lizards could be approached quite closely before they fled, but those with poor camouflage fled when the herpetologist was farther away. Insofar as flight reactions take over the burden of predator evasion, the body color is released to evolve with respect to temperature.

3. When an organism is adapted to a complete environment, and the different components begin to separate with climatic change, what strategy would be followed? Suppose for instance that a species is at its optimum temperature and substrate, but that secular change raises the temperature. It can move toward the new location of its optimum temperature, remain in its optimum substrate, select an intermediate habitat which is moderately satisfactory for both, or perhaps migrate so that the crucial stages of the life cycle have optimum temperature and substrate.

4. Since each part of the organism is the environment for other parts, we can study the evolution of internal diversity as the coevolution of parts and processes. Some examples of this are considered in the last chapter.

The theory outlined here suggests several lines of experiment and observation. First, it leads to direct predictions of a general sort expressed as inequalities comparing whole groups, such as the temperate and tropical faunae, tree and hole-nesting birds, or ant species with generalized and specialized caste polymorphisms.

A second line involves the detailed analysis of particular adaptive systems with the measurement of the fitness set, environmental uncertainty and information, and the actual strategy compared with the calculated optimum strategy. Diapause strategy in economically important insects or mites would be especially amenable to such study since there is widespread monitoring of their populations by agricultural experiment stations. We have already described a rough qualitative model. Further refinements for particular cases would lead to specific predictions as to the date of onset of diapause at different locations, population differences in the importance of different signals, and the uniformity of the switch from non-diapause to diapause forms. It would also lead to predictions as to the results of attempting to control the pests by confounding their information. If artificial lighting were used to delay the onset of diapause, light would cease to be a good predictor of season and the populations affected might become obligate diapausing univoltine races.

A third line of work is in the genetics of flexibility for different types of structures and processes. If the pattern of response is strategic, it would be selected separately for different aspects of the same organism instead of being controlled by some general "flexibility" factors, and may vary widely in the same character from group to group. The work of Bradshaw (1965) and his collaborators seems to support this view.

Finally, one could attempt to produce natural selection for an optimal strategy in laboratory populations. Since "strategy" is a phenotypic trait, there is no reason to doubt the possibility of such selection. The differentiation of the European corn borer (*Ostrinia nubilalis*) populations into distinct diapause races in the forty or so years since their introduction into North America encourages the belief that the selection may be strong enough to be rapidly effective. One such experiment would be to subject *Drosophila* larvae to a single day of high or low temperature in the last instar. The high-temperature treatment would be followed by selection of the first flies to emerge from the puparium. The low-temperature treatment would be followed by discarding the first flies and retaining the last group only. We would therefore expect natural selection for a pupal dormancy or pseudodiapause evoked by one day of later larval temperature. The next step from there would be to replace the temperature signal by light, or oil of wintergreen. We could then test the hypothesis of the arbitrary physical relation between the form of the signal and the evoked response, that almost any signal can be coupled to almost any responding system.

Beyond the specific experimental and observational testing of the theory, this approach directs attention toward the sufficient parameters of the environment, and would therefore direct bioclimatological work toward the definition and measurement of these evolutionarily significant characteristics of environment.

Finally, strategic analysis can be used to interpret the evolution of complex systems such as biochemical networks and behavioral modes, suggesting relevant measures and selection pressures.

The Theory of the Niche

The concept of the ecological niche has been used heuristically for a long time. Grinnell (1904) referred to the niches of birds to indicate that different species have different requirements. Elton (1927) used the niche mostly for defining a species' position in the trophic hierarchy. Biogeographers have long noted the apparent equivalence between organisms of different regions and corresponding numbers of species in corresponding formations. For instance, in Table 3.1 we show the percentages of bird species of the Maylayan and Australian rain forests in each level of the forest and for each food category. This correspondence suggests that the relative abundance of opportunities for species of each type, or the number niches, is the same in both forests. The notion of saturated environments versus those with empty niches has been used to compare insular and continental biotas. Finally, the classical experiments of Gause (1934) led to the exclusion principal: if two species coexist they must occupy different niches. But it was due to the work and inspiration of G. E. Hutchinson (1965) that progress has been made toward a rigorous theory of the niche.

A satisfactory theory of the niche must permit an accurate description of a species' niche, and must be able to use that description to answer the following questions:

1. What determines the degree of specialization of a species, or inversely, its niche breadth?
2. What determines the species diversity of a community in relation to area, climatic region, size of organism, trophic level, etc?
3. How similar can species be and yet coexist?
4. How do similar species divide the environment among them?