THEORETICAL POPULATION BIOLOGY 36, 125–143 (1989)

Evolutionary Stability in One-Parameter Models under Weak Selection

PETER D. TAYLOR

Department of Mathematics and Statistics,
Queen's University, Kingston, Ontario, Canada K7L 3N6

Received November 1, 1987

A general notion of evolutionary stability is formulated in models in which the possible behaviours are parameterized by a continuous variable, and selection is assumed to be weak. Two local stability conditions are formulated, m-stability and δ-stability, the former being first-order and the latter second-order in the mutant behavioural deviation. The conditions are interpreted in two standard formulations of a one-locus genetic model: a covariance approach and a structured population approach. A weak selection theorem is proved which says that m-stability can be calculated using the neutral covariances. These in turn can be calculated as relatedness coefficients; hence an inclusive fitness formulation is capable of checking m-stability. But δ-stability, being second-order, is more difficult to handle. © 1989 Academic Press, Inc.

1. INTRODUCTION

A general problem in the modeling of animal and plant behaviour is to find an evolutionarily stable configuration. The possible behaviours are indexed by a scalar or vector parameter m, and the objective is to find a value of m, or a mixture of such values, which is, in some sense, more fit than its alternatives. But since the fitness of any behaviour typically depends on the mix of behaviours present in the population, the definition of stability requires a careful formulation.

The idea behind the condition, that m^* be locally stable, is to consider a population which is a slight alteration, in some sense, from a pure m^*-population, and to require that the action of selection move the population back towards m^*. But this is not, in general, an easy condition to formulate: the number of types of alterations is often large, and since selective forces change as the population changes, it is not easy to perceive their long term effects.

0040-5809/89 $3.00
Copyright © 1989 by Academic Press, Inc.
All rights of reproduction in any form reserved.

There are two special cases that have received a lot of attention in the literature. The first is the classic ESS theory (Maynard Smith and Price 1973, Maynard Smith 1974) in which the fitnesses are assumed to be linear: the fitness of any strategy is a linear function of the population strategy mix. In this case, as far as fitnesses are concerned, many different kinds of alterations are equivalent, and the simple condition requiring m^* to be more fit than average in the altered population works well, at least locally (i.e., for small alterations). Here, m is typically a mixture of a finite number of "pure strategies," though in simple cases (e.g., the war of attrition) the number of strategies may be infinite.

The second special case is the one we are concerned with here. It relinquishes the linear structure of the fitness, but exploits a natural order structure on the parameter set. Suppose the possible m values lie along a line segment, as would be the case if m represented the probability of a certain action or the proportion of resources allocated to a fixed purpose. Then we can get a good feeling for the local action of selection in an m-population by simply examining the fitness of rare mutants which are near m on either side. In fact there are two natural conditions to look at in testing the local stability of m^*. One is to require that in an m^*-population, all local mutants be less fit, and the other is to take m near m^* and require that, in an m-population, local mutants on the m^*-side of m be more fit than those on the other side. I call the two types of stability conditions which result δ-stability and m-stability respectively. A formal analysis of these conditions was first done by Eshel and Motro (1981) and Eshel (1983).

In Section 2, I provide a formulation of the above two stability conditions in terms of a general fitness function, and in Sections 3 and 4, I interpret these conditions in a one-locus genetic model using, as the fitness function, the relative change in frequency of the mutant allele over one generation. In Section 3, I use a covariance approach and in Section 4, a structured population approach with the assumption of a rare mutant allele.

One difficulty with both these approaches lies in the calculation of the distribution of the mutant allele, for the deviant behaviour it occasions may alter its distribution from the more easily calculated neutral distribution. Theorem 2, the "weak selection" theorem, says that the neutral distribution will give the correct mutant fitness to first order in the behavioural deviation δ. We conclude that this neutral distribution will correctly verify the m-stability condition. A condition is given under which this distribution will suffice to check δ-stability as well, but in general this will not be the case. Theorem 3 summarizes the consequences of this result for the inclusive fitness approach.

2. Stability in a One-Parameter Model

I assume the range of possible behaviours is described by a continuous scalar parameter m which can be regarded as the probability of engaging in a certain activity, or the proportion of resources invested in one activity instead of another. I let m stand for "normal" behaviour, and I consider a "mutant" behaviour which deviates from m by an amount δ.

The objective of any model is to produce an expression $W(m, \delta)$ for the fitness of the mutant behaviour in the mixed population. It is convenient for us to let W represent the fitness *increment* of the mutant allele, that is, the average fitness difference between mutant and normal behaviour. Thus if $\delta = 0$, the mutant behaves normally, and so $W(m, 0) = 0$. In the genetic models of Sections 3 and 4, I will take W to be $\Delta Q/Q$, the relative change of frequency of the mutant allele over one generation.

The ESS Conditions

I now formulate the ESS conditions for a normal behaviour m^* to be stable to mutant invasion. I identify two types of local stability: to changes in normal parameter (m-stability) and to changes in mutant deviation (δ-stability). Under m-stability, if $m < m^*$, selection should favour mutants with $\delta > 0$, and if $m > m^*$, selection should favour mutants with $\delta < 0$. Under δ-stability, at $m = m^*$, all mutants should be less fit than normal. The local (δ small) formulations of these are:

Definition 1. The ESS conditions.

m^* is m-stable if for m near m^* and δ near 0,

whenever $m < m^*$, $W(m, \delta)$ has the same sign as δ, and
whenever $m > m^*$, $W(m, \delta)$ has the opposite sign of δ.

m^* is δ-stable if for δ near to but different from 0, $W(m^*, \delta) < 0$.

These conditions are illustrated geometrically in Fig. 1.

If W is differentiable, and m^* is an interior point in its range, each condition implies the *equilibrium condition*

$$\frac{\partial W}{\partial \delta}(m^*, 0) = 0 \tag{2.1}$$

whose solutions are called the *equilibrium points* of the model.

I now look at the Taylor series expansion of W about $\delta = 0$, and formulate the differential analogues of these two conditions. Recall that $W(m, 0) = 0$ for all m.

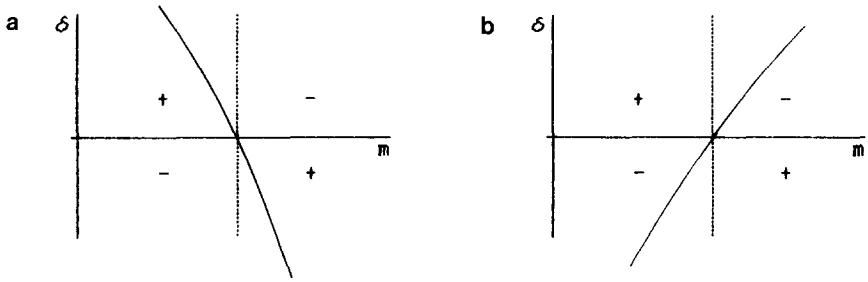

FIG 1. The sign of W near an m-stable point. Consider the set of points (m, δ) at which $W=0$. Since $W(m, 0)=0$, the m-axis is included in this set. For m^* to be an equilibrium point, there must be other curves in this set which cross the m-axis at m^*. In the simplest (and the generic) case, there will be one other such curve, and it, together with the m-axis, will divide the space around $(m^*, 0)$ into four regions. For m^* to be m-stable, the sign of W must be as shown in both diagrams. The point will be δ-stable if the vertical line at m^* is (locally) in the negative regions. This happens in 1(a) but not in 1(b).

THEOREM 1. *Suppose*

$$W(m, \delta) = \delta a(m) + \frac{\delta^2}{2} b(m) + o(\delta^2), \tag{2.2}$$

where $a = \partial W/\partial \delta$ *and* $b = \partial^2 W/\partial \delta^2$, *both evaluated at* $\delta = 0$.

(1) *If* $a(m) \neq 0$, *then for* δ *near* 0, *selection will favour mutants with* $\delta a(m) > 0$ *and disfavour mutants with* $\delta a(m) < 0$.

(2) *If* $a(m^*) = 0$, *then* m^* *is an equilibrium point and is*

$$m\text{-stable if } da/dm < 0 \text{ at } m = m^* \tag{2.3}$$

$$\delta\text{-stable if } b(m^*) < 0. \tag{2.4}$$

As second-order conditions in W, *these can be written*

$$m\text{-stable:} \quad \frac{\partial^2 W}{\partial m\, \partial \delta}(m^*, 0) < 0 \tag{2.3}'$$

$$\delta\text{-stable:} \quad \frac{\partial^2 W}{\partial \delta^2}(m^*, 0) < 0. \tag{2.4}'$$

An elementary analytic argument shows that (2.3) and (2.4) are strictly stronger than the formulations of Definition 1, though I would never expect to find a biological example where the definition held but the differential conditions failed.

Eshel and Motro (1981) and Eshel (1983) call (2.4) the ESS condition (evolutionarily stable strategy), and when both conditions hold the ESS is called continuously stable, or CSS, so named because (2.3) only makes sense when m is a continuous variable. My preference is to make both conditions part of what is generally called evolutionary stability for one-parameter models.

Fisher's Sex Ratio Argument

By way of illustration, I report on what happens when these definitions are applied to a couple of standard examples. The first example is Fisher's (1930) argument that an unbiased sex ratio is stable in a diploid population with random mating. He observed that in a population with a biased sex ratio, individuals who produced the rarer sex had an increased number of grandchildren, and if this tendency was inherited, the population sex ratio would move towards equal numbers of each sex. This is essentially an m-stability argument, because it looks at mutant fitness in a population in which normal behaviour produces a biased ratio. The question of δ-stability in this example concerns the mutant fitness in a population in which normal behaviour is unbiased. Mutant behaviour is also penalized here, but the effect is weak (of order Q) when the mutant is rare, because the disadvantage to the mutant is proportional to the deviation of the population mean ratio from 1/2. This was first pointed out by Shaw and Mohler (1953).

Matrix Games

My second example is the classical ESS theory of matrix games (Maynard Smith and Price 1973, Maynard Smith 1974) with the possibility of contests between relatives, treated by Grafen (1979) and Hines and Maynard Smith (1979). If there are two pure strategies, in a matrix game, the set of possible mixed strategies is a line segment, and the general analysis of this paper should apply. Since the standard ESS definition would seem to provide a complete condition for stability, one is curious to know how this definition relates to the two conditions of this paper. In fact they turn out to be all equivalent. If m is the probability of playing strategy 1, and the payoff matrix is $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, then the equilibrium strategy is (Grafen 1979)

$$m = -\frac{b - d + R(c - d)}{(1 + R)(a - b - c + d)},$$

where R is the average relatedness of a player to his opponents. If this is between 0 and 1, the conditions for m- and δ-stability both reduce to the standard ESS condition (Grafen 1979) that $a - b - c + d$ be negative.

Roughly speaking, because of the linear form of the fitness function, the effects on mutant fitness of shifting m from equilibrium, and of shifting δ from 0, are the same.

Comparison of the Two Conditions

I have analyzed the two stability conditions in a genetic model of seed provisioning (Taylor 1989), and have constructed an example in which δ-stability holds but not m-stability. A mathematical example of this type was given by Eshel and Motro (1981). It is also easy to provide *mathematical* examples of functions with equilibrium points which are m-stable but not δ-stable, but I have yet to see a *biologically* plausible example of this. It is worth asking why this might be so, especially since, as we shall see below, m-stability is much easier to verify than δ-stability. Let me offer a rather vague reason.

As we can see from Fig. 1, from a *mathematical* point of view, m-stable points should equally likely be δ-stable or not, depending on whether the line $W=0$ crosses the m-axis from the left (Fig. 1a) or from the right (Fig. 1b). Why, in *biological* examples, should it be more likely to cross from the left? Often, in natural examples, such as the sex ratio example above, the fitness of a rare mutant depends not so much on m as on the population average $\bar{m}=(1-Q)\,m+Q\delta$. That is, for small Q and perhaps small δ, the sign of $W(m, \delta)$ depends mainly on the sign of δ and the value of $\bar{m}$. Thus, on each side of the m-axis, W has a constant sign along lines of constant $\bar{m}$. For small Q, this gives us a family of lines of steep negative slope $-1/Q$, which the line $W=0$ should tend to be parallel to, at least for small δ. This argues for Fig. 1(a).

3. The Genetic Model: A Covariance Approach

I suppose there are two alleles at a single locus, a normal allele and a mutant allele with frequency Q, and I let ΔQ denote the change in Q over one generation of selection. I let δ measure the relative effect (in a manner to be specified) of the mutant allele on individual behaviour.

In this section I obtain an expression for ΔQ in terms of the covariances between genotype and phenotype. These covariances are generally hard to calculate, because the action of the mutant allele affects its distribution, and this distribution is hard to find. What is often easy to find is the *neutral* distribution that obtains when $\delta=0$. If this neutral distribution is used to calculate the change in frequency of the mutant allele, we get an approximation of ΔQ which I denote by $\Delta^0 Q$. Thus, $\Delta^0 Q$ is typically much easier to calculate than ΔQ. Theorem 2 tells us that if the mutant allele is rare, ΔQ and $\Delta^0 Q$ are equal to first order in δ. The important consequence

of this is that $\Delta^0 Q$ can be used to verify m-stability. In Theorem 3, I relate this result to the inclusive fitness approach.

In the following analysis, both mutant frequency Q and mutant behavioural deviation δ may be assumed to be small, and to formulate the results clearly it is useful to use the "big O" and "little o" notation, whose meanings I recall. A quantity K is "big O of α", written $O(\alpha)$, if K approaches zero as α approaches zero, and a quantity K is "little o of α" written $o(\alpha)$, if K/α approaches zero as α approaches zero.

I assume the population is homogeneous and consists of individuals all of the same type. In Taylor (1988b) I have discussed the extension of these results to cases in which both males and females are involved in the action, but in different ways. I will let y and z denote random *actors*, that is, individuals whose behaviour affects the fitness of others, and I let x denote a random *recipient*, that is, an individual whose fitness is affected by the mutant behaviour.

Change of Allele Frequency

We can get an expression for ΔQ from a covariance formula of Price (1970),

$$\Delta Q = \operatorname{cov}(G_x, w_x)/\bar{w}, \tag{3.1}$$

where x is a random member of the population, G_x is the genotypic value of x, defined as the frequency in x of the mutant allele, w_x is the fitness of x, and $\bar{w}$ is mean fitness. This expression assumes Mendelian assortment of gametes to offspring, that is, any allele donated by x to an offspring will be mutant with probability G_x.

Now I obtain an expression for w_x. The idea is that the fitness of each individual x will depend on the behaviour of every individual y in the population (including x himself), and this behaviour will, in turn, depend on genotype. I allow individuals to adopt different levels of mutant behaviour, and let

$$m_y = m + H_y \delta \tag{3.2}$$

be the behaviour practised by y. H_y is called the phenotypic value of y, and will be determined by his genotype, or, more generally, by the genotype of the individual controlling his behaviour (Taylor 1988a). A standard diploid model takes $H_y = G_y$ when $G_y = 0$ or 1, and $H_y = h$ when $G_y = 1/2$.

Now I define

$$\partial w_x/\partial m_y = s_{xy} \tag{3.3}$$

$$\partial^2 w_x/\partial m_y\, \partial m_z = s_{xyz} \tag{3.4}$$

to be the first- and second-order effects of the behaviour of y (or of y and z) on the fitness of x (Hamilton 1970), where all derivatives are evaluated at $\delta = 0$. Then, expanding in powers of δ,

$$w_x = w_0 + \delta \sum_y H_y s_{xy} + \frac{\delta^2}{2} \sum_y \sum_z H_y H_z s_{xyz} + o(\delta^2), \tag{3.5}$$

where w_0 is fitness in a normal population, and the sums are over the whole population, including x himself. I note that average fitness is

$$\bar{w} = w_0 + \frac{\delta}{N} \sum_y H_y s_y + o(\delta), \tag{3.6}$$

where $s_y = \sum_x s_{xy}$ is the sum of the effects of y on members of the population. Since the average phenotype $\sum H_y/N$ is of order Q, Eq. (3.6) can be written

$$\bar{w} = w_0 + \delta O(Q) + o(\delta), \tag{3.7}$$

which we invert to get

$$(\bar{w})^{-1} = (w_0)^{-1} + \delta O(Q) + o(\delta). \tag{3.8}$$

If we put (3.5) and (3.8) into Price's covariance formula (3.1), we get

$$\begin{aligned}
\Delta Q = & \left[\frac{1}{w_0} + \delta O(Q) + o(\delta) \right] \\
& \times \left[\delta \sum_y \operatorname{cov}(G_x, H_y s_{xy}) + \frac{\delta^2}{2} \sum_y \sum_z \operatorname{cov}(G_x, H_y H_z s_{xyz}) + o(\delta^2) \right] \\
= & \frac{1}{w_0} \left[\delta \sum_y \operatorname{cov}(G_x, H_y s_{xy}) + \frac{\delta^2}{2} \sum_y \sum_z \operatorname{cov}(G_x, H_y H_z s_{xyz}) \right] \\
& + \delta^2 o(Q) + o(\delta^2) \\
= & \frac{1}{w_0} \left[\delta \sum_i n_i \operatorname{cov}_i(G_x, H_y) s_i + \frac{\delta^2}{2} \sum_j n_j \operatorname{cov}_j(G_x, H_y H_z) s_j^{(2)} \right] \\
& + \delta^2 o(Q) + o(\delta^2),
\end{aligned} \tag{3.9}$$

where I have grouped interactions with the same average effect, with s_i and $s_j^{(2)}$ being the different first- and second-order effects, and n_i and n_j the number of interactions of a random x with these effects. Here, cov_i is over all x–y pairs with effect s_i and cov_j is over all x–y–z triples with effect $s_j^{(2)}$. The $O(Q)$ in the first expression has become $o(Q)$ in the second because the cov_i terms are $O(Q)$.

Appearances to the contrary, (3.9) is not yet the second-order expansion of ΔQ in δ, for the covariances will usually depend on δ: the deviant behaviour of the mutant allele may alter its distribution, and hence alter the covariances. The trouble is that this altered distribution can be difficult to calculate. Of course, all we need for the second-order expansion of ΔQ is the δ term of the expansion of cov_i (see (3.11) below), but this may not be easy to calculate. However, what *are* often easy to calculate are the neutral covariances which belong to the $\delta = 0$ distribution, and what we now do is calculate the approximation $\Delta^0 Q$ obtained by using this distribution, and see how good it is. From (3.9)

$$w_0 \Delta^0 Q = \delta \sum_i n_i \mathrm{cov}_i^0(G_x, H_y) s_i + \frac{\delta^2}{2} \sum_j n_j \mathrm{cov}_j^0(G_x, H_y H_z) s_j^{(2)} + \delta^2 o(Q) + o(\delta^2), \tag{3.10}$$

where I have used the superscript ^{0} on the covariances to signal the use of the neutral distribution. The question is, when will $\Delta^0 Q$ give us the correct local stability analysis using the conditions of Theorem 1? If I define c_i by the equation

$$\mathrm{cov}_i(G_x, H_y) = \mathrm{cov}_i^0(G_x, H_y) + \delta c_i + o(\delta) \tag{3.11}$$

then (3.9) can be written

$$w_0 \Delta Q = \delta \sum_i n_i \mathrm{cov}_i^0(G_x, H_y) s_i + \frac{\delta^2}{2} \left[\sum_j n_j \mathrm{cov}_j^0(G_x, H_y H_z) s_j^{(2)} + 2 \sum_i n_i c_i s_i \right] + \delta^2 o(Q) + o(\delta^2), \tag{3.12}$$

and comparing (3.10) and (3.12) we have

$$\frac{\Delta Q}{Q} - \frac{\Delta^0 Q}{Q} = \frac{\delta^2}{Q} \left[\sum_i n_i c_i s_i + o(Q) \right] + o(\delta^2). \tag{3.13}$$

The results are summarized in the following theorem.

THEOREM 2. (The weak selection theorem). *Suppose there are two alleles at one locus, one for the normal value m of the behavioural parameter, and the other, of frequency Q, for the mutant value, with deviation δ. Suppose that in the covariance formula* (3.1) *for ΔQ, the neutral ($\delta = 0$) distribution of the mutant allele is used, and denote by $\Delta^0 Q$ the resulting approximation.*

(1) *$\Delta Q/Q$ and $\Delta^0 Q/Q$ agree to first order in δ. Thus, away from equilibrium, and for δ near 0, $\Delta^0 Q/Q$ will tell us whether the mutant allele is increasing or decreasing in frequency. It follows that $\Delta^0 Q/Q$ will identify all equilibrium points, and will correctly predict their m-stability.*

(2) *If for all interactants x and y, $\mathrm{cov}(G_x, H_y)$ is independent of δ (such that all $c_i = 0$), or, more generally, if the c_i in (3.11) are $o(Q)$, then the δ^2 terms of $\Delta Q/Q$ and $\Delta^0 Q/Q$ will agree to order $O(Q)$ and, for sufficiently small mutant frequency, $\Delta^0 Q/Q$ will correctly predict δ-stability.*

An important case in which the condition of (2) holds is a diploid population with outbreeding in which interactions are only between members of the same nuclear family (between sibs or between parent and offspring), for then there is, to order Q, only one type of mutant family, with one mutant and three normal parental alleles. Ignoring mutant families with more than one parental mutant allele, which will occur with frequency $o(Q)$, the resulting distribution of mutant alleles among members of the same family is independent of δ, and hence the c_i in (3.11) are $o(Q)$. An example of this is found in the seed-provisioning model of Queller (1984) discussed by Taylor (1989), where the interactions are actually between half-sibs. However, with inbreeding, the c_i term in (3.11) must be reckoned with in the δ^2 analysis of ΔQ.

The Inclusive Fitness Approach

I now show that the first-order term of ΔQ can be obtained with an inclusive fitness calculation. Fasten attention on a fixed actor y. From (3.9), he can expect n_i interactions of type s_i among all individuals x whose fitness his behaviour affects (and y himself may be such an x). Thus the first sum in (3.12) may be considered to run over all such x, and

$$w_0 \Delta Q = \mathrm{cov}^0(G_y, H_y) w_I + o(\delta), \tag{3.14}$$

where

$$w_I = \sum_x R_{xy} s_{xy} \delta \tag{3.15}$$

is the inclusive fitness effect of y and

$$R_{xy} = \frac{\mathrm{cov}^0(G_x, H_y)}{\mathrm{cov}^0(G_y, H_y)} \tag{3.16}$$

is the relatedness between x and y. To get w_I we need to calculate the R_{xy} and these are often easily found. If H_y depends on G_y and y is outbred or the dependence is affine (additive gene action), then R_{xy} can be calculated with coefficients of consanguinity (Michod and Hamilton 1980, Pamilo and Crozier 1982, Grafen 1985, and Taylor 1988b) and so (3.14) is an important representation of ΔQ. The situation is summarized in the following theorem.

THEOREM 3. (The inclusive fitness theorem). *Assuming* $\mathrm{cov}^0(G_y, H_y)$ *is positive, the inclusive fitness* w_I *can be used to measure the effect of selection on the mutant allele, as follows*:

(1) *The mutant frequency will increase if* w_I *is positive and decrease if* w_I *is negative.*

(2) *If* $w_I = 0$, m^* *is an equilibrium point which is m-stable if* dw_I/dm *is negative.*

Thus, the usual inclusive fitness analysis can identify the equilibrium points and check their m-stability, but not their δ-stability. Note that Theorem 3 is a first-order result (in δ) and does not require the mutant allele to be rare.

4. THE GENETIC MODEL: A STRUCTURED POPULATION APPROACH

An important class of models concerns a population organized into discrete patches within which interactions are at random. As an example to keep in mind, consider a patch to be the set of all offspring of a group of N mated females. The analysis keeps track of the mutant allele by classifying patches according to "mutant type," determined by the distribution of mutant alleles in the patch. This structured population approach has its own methods and notation, and my purpose here is to relate these to the covariance approach of the last section. In this section, I assume the mutant allele is rare; this provides us with a linear transition from one generation to the next (Eq. (4.3)).

Let the mutant patch type be determined by the number of individuals of each mutant genotype on the patch. For example, for patches of sibs, if the number of offspring is large, the patches can be classified by parental (maternal and paternal) mutant genotype. I use the index k to keep track of different mutant patch types. I let u_k be the frequency of patch type k, and q_k be the frequency of the mutant allele in a type k patch. Then the overall mutant frequency is

$$Q = q \cdot u = \sum_k q_k u_k, \tag{4.1}$$

where the dot signifies the scalar product between the vectors $q = (q_k)$ and $u = (u_k)$.

The covariance approach of Section 3 requires the calculation of $\mathrm{cov}(G_x, H_y)$, where x and y are random individuals from the same patch, and, though I will not make use of this, it is interesting to see how this can be obtained from the above concepts. Within each patch, $\mathrm{cov}(G_x, H_y) = 0$,

since x and y are independent of one another, and so the overall covariance is just the covariance of patch means across patches. The mean of G_x over a type k patch is just q_k and I let H_k denote the mean of H_y. Then

$$\operatorname{cov}(G_x, H_y) = \sum_k u_k(q_k - \bar{q}_k)(H_k - \bar{H}_k),$$

where $\bar{q}_k$ and $\bar{H}_k$ denote the means of q_k and H_k over the population. Now $\bar{q}_k = Q$, and so

$$\operatorname{cov}(G_x, H_y) = \sum_k u_k(q_k - Q)H_k. \tag{4.2}$$

To calculate this we need to know the patch distribution vector u and I now turn to this question.

Of course u may change from generation to generation, and I denote by u' the distribution after one generation. Then u' will depend on u, and I assume this dependence is differentiable. If the mutant is rare the dependence will be nearly linear; more precisely,

$$u' = Au + o(Q), \tag{4.3}$$

where $A = A(m, \delta)$ is the patch type transition matrix. In fact (4.3) is the linear term of the Taylor series of u' as a function of u (when $u = 0$, $u' = 0$) and the entries of A are $a_{kh} = \partial u'_k / \partial u_h$, the number of extra type k patches next generation created by an extra type h patch this generation.

Now note that, with error $o(Q)$,

$$\Delta Q = q \cdot (u' - u) = q[A(m, \delta) - I]u \tag{4.4}$$

for any u. If $\delta = 0$, the mutant allele is neutral and should not change in frequency, no matter what its distribution, and so $\Delta Q = 0$ for all u. We deduce that

$$q[A(m, 0) - I] = 0 \tag{4.5}$$

and hence that q is a left eigenvector of $A(m, 0)$ for the eigenvalue $\lambda = 1$. This will be important in a moment.

Equation (4.3) implies that, with an error of $o(Q)$, the mutant allele distribution is multiplied by A each generation, and eventually u will converge to the dominant right eigenvector of $A(m, \delta)$, which I call $u(m, \delta)$, normalized so that $q \cdot u(m, \delta) = Q$. From this point on, the effect of each generation of selection will be to multiply u by the dominant eigenvalue $\lambda(m, \delta)$ of A, and (4.4) becomes, with error $o(Q)$,

$$\Delta Q = q[\lambda(m, \delta) - 1)]\, u(m, \delta) = (\lambda - 1)Q. \tag{4.6}$$

This is the same ΔQ that we calculated in (3.9), and this formula is difficult to work with for just the same reason as before: the eigenvalue $\lambda(m, \delta)$ is hard to calculate, and this is because the asymptotic distribution $u(m, \delta)$ is difficult to find. However, we get the same neutral approximation result as before. The eigenvector $u(m, 0)$, which is the asymptotic distribution of a neutral mutant, has eigenvalue $\lambda(m, 0) = 1$, and for this reason is often easy to calculate. If we use this instead of $u(m, \delta)$ in (4.4), what we get is exactly what we have called $\Delta^0 Q$, and so the analogue of (3.10) is, with error $o(Q)$,

$$\Delta^0 Q = q[A(m, \delta) - I]\, u(m, 0). \tag{4.7}$$

Now to relate (4.4) and (4.7), we differentiate (4.4) to get, with error $o(Q)$,

$$\frac{\partial \Delta Q}{\partial \delta} = q \frac{\partial A}{\partial \delta} u + q(A - I) \frac{\partial u}{\partial \delta} \tag{4.8}$$

and at $\delta = 0$,

$$\frac{\partial \Delta Q}{\partial \delta} = q \frac{\partial A}{\partial \delta} u = \frac{\partial \Delta^0 Q}{\partial \delta}, \tag{4.9}$$

using (4.5). This gives us another proof of Theorem 2 (1) for the case of a rare mutant.

If we differentiate (4.8) again we find that, at $\delta = 0$,

$$\frac{\delta^2 \Delta Q}{\partial \delta^2} = \frac{\partial^2 \Delta^0 Q}{\partial \delta^2} + 2q \frac{\partial A}{\partial \delta} \frac{\partial u}{\partial \delta} \tag{4.10}$$

with error $o(Q)$. From the δ^2 term of (3.10) and (3.12), we see that the last term of (4.10) is an alternative expression for $2 \sum n_i c_i s_i / w_0$ and, as before, its existence is the reason that $\Delta^0 Q$ is unable to analyze δ-stability. If the distribution of the mutant allele is independent of δ, then $\partial u / \partial \delta = 0$, and the last term of (4.10) vanishes, giving us an alternative proof of Theorem 2 (2).

8. Discussion

The Two Stability Conditions

Theorem 1 presents the differential forms (2.3) and (2.4) of the two stability conditions, and we discover an important mathematical difference between them in terms of the function $W(m, \delta)$: m-stability is first-order in

δ, and δ-stability is second-order. The practical consequence of this in any particular analysis is that m-stability is a lot easier to verify than δ-stability, because the first-order coefficient is typically the easier to calculate. Part of the reason for this has to do with the difficulty of finding out the distribution of the mutant allele when $\delta \neq 0$ and is discussed in Sections 3 and 4.

Conditions (2.3) and (2.4) are certainly not equivalent, and one can easily find functions $W(m, \delta)$ with equilibrium points at which each holds without the other. Whether one can find biologically interesting examples of this type is another question. In all the examples I have looked at, m-stable equilibrium points have always turned out to be δ-stable. Since δ-stability is the harder of the two conditions to verify, a result which provided general conditions under which m-stability would imply δ-stability would be of some practical significance. This point is discussed at the end of Section 2.

One might ask what happens to equilibria which are δ-stable but not m-stable. The answer is that, although at the exact equilibrium point all rare local mutants are penalized, the overall population m value should drift sideways (by some unspecified genetic or environmental mechanism), and then mutants which take it farther away will be favoured. The end result will be the establishment either of a stable equilibrium at another point, or of a polymorphic equilibrium which might straddle the original point. On the other hand, at an m-stable equilibrium which is not δ-stable, selection pressure will keep m from drifting, but will allow the spread of mutants with $\delta > 0$ or $\delta < 0$ or both. This results in a polymorphic population which is not described by the function $W(m, \delta)$.

Extension to Polymorphic Equilibria

There should be a natural extension of the stability conditions to equilibria in which there are two or more values of m present in the population. For example, suppose there are two normal phenotypes m_1 and m_2 of frequency p_1 and p_2, respectively, controlled at a single locus. First of all, the p_i must be at a stable "ecological" equilibrium, and the condition for this will typically depend on the underlying genetics. Given this, we want to form conditions for the evolutionary stability of the m_i. In the simple case in which the phenotypes are determined by two alleles a and A, with A dominant, we consider mutant alleles which are mutant forms of either A or a, and, in each case, only one of the m_i will be altered at a time, and the other can be held fixed. For each i, the fitness $W(m_i, \delta)$ is a function of only two variables, as before, and the local stability conditions are formed as in (2.3) and (2.4). But there are still difficulties in calculating $\mathrm{cov}(G_x, H_y)$ in (3.9) because H_y will depend, not only on the frequency G_y of the mutant allele in y, but on which non-mutant alleles happen to be

present, and this will be correlated with x's genotype. So different mutant genotypes must be considered separately, and this can lead to difficulties, and may require the population structure approach of Section 4.

Extension to Two-Parameter Models

Suppose we have a two-parameter behavioural space, for example sex ratio r, and dispersal probability m, assumed to be controlled at two loci, and the fitness of a female depends on both her parameters. A monomorphic equilibrium is found by treating each parameter separately, and solving two equations of the form (2.2) in the two unknowns r and m. But the stability analysis will depend on what deviations we want to allow. If we are prepared to assume that mutants are rare enough, spatially or temporally, that mutants at one locus will never encounter mutants at the other, then we simply have two one-parameter problems. But otherwise, we must allow for the possibility that genes at the two loci will assort non-randomly (linkage disequilibrium), and because of the fitness interactions, this will create a problem in the analysis.

The Genetic Models

In Section 3 and 4, using two different approaches, I interpret the stability conditions in a one-locus genetic model in which fitness is taken to be the change of frequency ΔQ over one generation of a rare mutant allele. The importance of the covariance approach of Section 3 is that the calculations are often simpler, and the components in the expressions usually have biological significance. On the other hand, the population structure approach of Section 4 often provides the only way of finding out the mutant distribution with enough precision to check the δ-stability.

Theorem 2 tells us that the first-order (in δ) term of ΔQ can be calculated using the neutral distribution of the mutant allele. This is an important result for both approaches, because calculations are much easier with this distribution. This is especially true in the covariance approach of Section 3, where the neutral covariances in the δ term of (3.12) can be calculated as relatedness coefficients. But the second-order term is not in general so easy to calculate, and the difficulties are displayed in the two parts of the δ^2 term of (3.12): in the first sum, the covariance involves quadratic terms in the phenotypic values, and in the second sum, we require knowledge of the actual (non-neutral) mutant distribution.

In Section 4, with the assumption of a rare mutant, the distribution of the mutant allele is obtained as the dominant right eigenvector of the transition matrix A, and this gives us an alternative approach to Theorem 2. In the neutral case, the corresponding eigenvalue is $\lambda = 1$, and the eigenvector is typically easy to obtain, but if $\delta \neq 0$ the calculation is often intractable, especially if A is of high dimension. As an example, if the patch type is

determined by the genotype of N mated females, the dimension of A can grow quickly with N: if $N=1$, there are 5 mutant patch types, and if $N=2$, this number rises to 20. Examples of this approach can be found in Uyenoyama and Bengtsson (1981) and Uyenoyama (1984).

The Inclusive Fitness Analysis

The calculation of the neutral covariances which appear in the first-order term of (3.10) is most elegantly done using relatedness coefficients, and this leads to the inclusive fitness formulation (3.14) of the first-order term of ΔQ. A number of recent papers (Hamilton 1979, Michod 1979, Charlesworth 1980, Michod and Hamilton 1980, Seger 1981, Uyenoyama and Feldman 1981, Michod 1982, Karlin and Matessi 1983, Queller 1984, Uyenoyama 1984, Grafen 1985, and Taylor 1988b) have discussed the relationship between genetic and inclusive fitness models, mainly in the context of altruistic behaviour, and this paper builds on these efforts. Theorem 3 gives a precise statement of the extent to which inclusive fitness can be used to measure the change in frequency of the mutant allele, and in particular shows that w_1 can be used to determine m-stability. Two things must be emphasized: first that m-stability is a local condition (in δ) and Theorem 3 will only be of practical consequence in models in which selection is weak. The second is that the other ESS condition, δ-stability, cannot be tested with the usual inclusive fitness formulation. This has been pointed out by a number of authors, and examples of this in models of altruism are found in Uyenoyama (1984).

Care must be taken with the inclusive fitness formulation if fitness interactions involve individuals of both sexes, and there are parental asymmetries in the passing of genes to offspring. Then the relatednesses must be weighted with reproductive values. This question has been discussed in Taylor (1988b, 1990).

Assumptions of Additivity

This equivalence between genetic and inclusive fitness models is at first puzzling. Inclusive fitness arguments are often so simple because they simply add up the fitness effects (weighted by relatedness coefficients) of a single mutant individual, but this can only be valid if fitness effects between individuals are additive. Where is the assumption of additivity in the above result? The answer is that this is contained in the assumption that the w_x are differentiable functions of the m_y, for differentiable functions are always additive to first order, and the equivalence only holds to first order in the behavioural deviations. Models in which fitness interactions are non-additive, such as warning call behaviour (perhaps one caller is as good as two), can be converted to models of this type by thinking in terms of a continuous variable such as the probability m of making the call. Then, in

an m-population with some mutant behaviour, the probability of getting a call in a certain group will, to first, depend additively on the mutant deviations.

There are two other places in the above development in which assumptions of additivity arise. Price's covariance formula, in the form given above, requires average offspring genotypic value to depend additively on parental genotypic value, and this is effected by my assumption of Mendelian assortment of alleles. Also, the relatedness coefficient (3.16) is only given by the easily computable pedigree coefficient of relationship (Pamilo and Crozier 1982), defined in terms of identity of genes by descent, when H_y in (3.16) can be replaced by G_y, and one way to make this happen is to assume H_y depends linearly on G_y. This is essentially an assumption of additive gene action within an individual. Michod and Hamilton (1980) and Taylor (1988b) have further discussions of this matter.

Terminology

There are some matters of notation and terminology that I find perplexing. One has to do with the notions of m- and δ-stability. They seem to me to describe the two aspects of stability that I think are fundamental in these one-parameter models, but I am not sure what to call them. The problem with my terminology is that it involves the names of the variables. I have considered the terms "normal-stability" and "mutant-stability" but I am not happy with these.

Another issue has to do with whether the actor A or the recipient B should come first in subscript notation. This question arises not only for relatedness coefficients, but for fitness effects, and I feel the same convention should govern each. The classical literature lists the actor first. Thus Hamilton's (1964, 1972) relatedness coefficients are written r_{AB} or b_{AB} and his (1970) "effects" of A's action on B's fitness are denoted s_{AB}. And there is a sense in which this is natural, and corresponds to the way we think and talk: when an actor has an effect, then the act comes first and the effect comes second. Also, s_{AB} is naturally described as the effect of A on B. In Taylor (1988b), I followed the notation of Uyenoyama (1984) and wrote my relatedness coefficients as $R_{A\to B}$ with the arrow emphasizing the effect of A on B, and I wrote my effects as s_{AB}.

But there are good arguments for the opposite convention. Pamilo and Crozier (1982) note that since the relatedness coefficient is, in simple cases, the coefficient of regression of G_B on G_A [this happens when gene action is additive or when A is outbred (Michod and Hamilton 1980, Grafen 1985)], the regression notation which puts B first should be followed. And considering effects, if the effect of A on B is defined, as in (3.4), as $\partial w_B/\partial m_A$ then the "matrix" of effects is the Jacobian matrix of the vector function

$w = (w_x)$ of $m = (m_y)$, and the entries of this matrix are conventionally written with the dependent variable indexing the rows and the independent variable the columns. This suggests that B should come first. In any case, it is this latter scheme that I have adopted in this paper. It would be nice to see some general agreement for one convention or the other.

ACKNOWLEDGMENTS

I am grateful to Marcy Uyenoyama for her painstaking criticisms of numerous drafts of this paper and of much of the work on which it is based. David Queller, Alan Grafen, Michael Bulmer, Steve Frank, Ilan Eshel, and the referees have also provided valuable comments. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

CHARLESWORTH, B. 1980. Models of kin selection, in "Evolution of Social Behaviour: Hypotheses and Empirical tests" (H. Markl, Ed.), Verlag Chemie, Weinheim.

ESHEL, I. 1983. Evolutionary and continuous stability, *J. Theor. Biol.* **103**, 99–111.

ESHEL, I., AND MOTRO, U. 1981. Kin selection and strong evolutionary stability of mutual help, *Theor. Pop. Biol.* **19**, 420–433.

FISHER, R. A. 1930. "The Genetical Theory of Natural Selection," Claredon Press, Oxford (Reprinted and revised 1958, 1981, Dover, New York).

GRAFEN, A. 1979. The hawk dove game played between relatives, *Anim. Behav.* **27**, 905–907.

GRAFEN, A. 1985. A geometric view of relatedness, *Oxford Surv. Evol. Biol.* **2**, 28–89.

HAMILTON, W. D. 1964. The genetical evolution of social behaviour, I and II, *J. Theor. Biol.* **7**, 1–52.

HAMILTON, W. D. 1970. Selfish and spiteful behaviour in an evolutionary model, *Nature* **228**, 1218–1220.

HAMILTON, W. D. 1972. Altruism and related phenomena, mainly in social insects, *Annu. Rev. Ecol. Syst.* **3**, 192–232.

HAMILTON, W. D. 1979. Wingless and fighting males in fig wasps and other insects, in "Reproductive Competition and Sexual Selection in Insects" (M. S. Blum and N. A. Blum, Eds.), Academic Press, New York.

HINES, W. G. S., AND MAYNARD SMITH, J. 1979. Games between relatives, *J. Theor. Biol.* **79**, 19–30.

KARLIN, S., AND MATESSI, C. 1983. Kin selection and altruism, *Proc. Royal Soc. London Ser. B* **219**, 327 353.

MAYNARD SMITH, J. 1974. The theory of games and the evolution of animal conflicts, *J. Theor. Biol.* **47**, 209–221.

MAYNARD SMITH, J., AND PRICE, G. R. 1973. The logic of animal conflict, *Nature* **246**, 15–18.

MICHOD, R. E. 1979. Genetical aspects of kin selection: effects of inbreeding, *J. Theor. Biol.* **81**, 223–233.

MICHOD, R. E. 1982. The theory of kin selection, *Annu. Rev. Ecol. Syst.* **13**, 23–55.

MICHOD, R. E., AND HAMILTON, W. D. 1980. Coefficients of relatedness in sociobiology, *Nature* **288**, 694–697.

PAMILO, P., AND CROZIER, R. H. 1982. Measuring genetic relatedness in natural populations: methodology, *Theor. Pop. Biol.* **21**, 171–193.

PRICE, G. R. 1970. Selection and covariance, *Nature* **227**, 520–521.

QUELLER, D. C. 1984. Models of kin selection in seed provisioning, *Heredity* **53**, 151–165.

SEGER, J. 1981. Kinship and covariance, *J. Theor. Biol.* **91**, 191–213.

SHAW, R. F., AND MOHLER, J. D. 1953. The selective significance of the sex ratio, *Amer. Nat.* **87**, 337–342.

TAYLOR, P. D. 1988a. An inclusive fitness model for dispersal of offspring, *J. Theor. Biol.* **130**, 363–378.

TAYLOR, P. D. 1988b. Inclusive fitness models with two sexes. *Theor. Pop. Biol.* **34**, 145–168.

TAYLOR, P. D. 1989. Local stability in a seed provisioning model. *J. Theor. Biol.*, in press.

TAYLOR, P. D. 1990. Allele frequency change in a class-structured population. *Amer. Nat.*, in press.

UYENOYAMA, M. K. 1984. Inbreeding and the evolution of altruism under kin selection: effects on relatedness and group structure, *Evolution* **38**, 778–795.

UYENOYAMA, M. K., AND BENGTSSON, B. O. 1981. Towards a genetic theory for the evolution of sex ratio, II. *Theor. Pop. Biol.* **20**, 57–79.

UYENOYAMA, M. K., AND FELDMAN, M. W. 1981. On relatedness and adaptive topography in kin selection, *Theor. Pop. Biol.* **19**, 87–123.