LINKAGE AND SELECTION: THEORETICAL ANALYSIS OF THE DETERMINISTIC TWO LOCUS RANDOM MATING MODEL

WALTER F. BODMER[1] AND JOSEPH FELSENSTEIN

*Department of Genetics, Stanford University School of Medicine, Palo Alto, California[2] 94304
and Department of Zoology, University of Chicago, Chicago, Illinois 60637*

CONTENTS

A simple model for the evolution of closer linkage between two linked loci was outlined by R. A. FISHER in 1930. Though unaccompanied by any theoretical analysis FISHER's brief discussion was the basis for most of the subsequent analysis of the interaction between selection and linkage. The first attempt at a specific analysis of the two locus case was by WRIGHT (1952) using a simple symmetric viability model. KIMURA (1956) analyzed a model which showed how natural selection could give rise to closer linkage between two loci. LEWONTIN and KOJIMA (1960) later derived equilibrium frequencies and stability conditions for some further special two locus cases which illustrated the importance of the interaction between linkage and viability interactions. The interaction between selection and linkage in evolution was extensively reviewed by BODMER and PARSONS (1962). Their analysis of the two locus case placed special emphasis on the effects of linkage on the conditions for the increase of newly arisen gene complexes. Many papers have appeared since this review which extend these earlier results in various directions. In particular there has been some emphasis on more

penetrating numerical investigations of the two locus case (KOJIMA 1965; LEWON-TIN 1964 a,b,c; PARSONS 1963 a,b; JAIN and ALLARD 1965; SINGH and LEWONTIN 1966) and on evaluating the effects of inbreeding on the interaction between selection and linkage (JAIN and ALLARD 1966). The main aim of this paper is to provide the theoretical background for some of the results given by BODMER (in BODMER and PARSONS 1962) and to extend and amplify the treatment of the two locus case with random mating. No specific attempt will be made to review all of the more recent literature though we have endeavored to include most of the relevant papers in the list of references.

After reviewing the derivation of the general equations we obtain a necessary condition on the viabilities for the existence of a "linkage equilibrium" and show that this is satisfied by simple multiplicative and additive models. We determine next the nature of the equilibria given by a general symmetrical model and then examine the conditions for the increase of newly produced gametic combinations. Finally we shall make some attempt to synthesize the overall properties of the two locus, random mating equations.

1. *The General Model*

In a random mating diploid population with discrete generations, the fitness of any genotype will be proportional to the probability of its survival to maturity multiplied by its fertility. BODMER (1965) has shown that if the fertility of a mating is the product of the fertility values of the parents, Hardy-Weinberg proportions will obtain in the offspring provided that the fertility values of the sexes are the same. These conditions are assumed to hold in all the following derivations. In addition, we ignore random fluctuations in gene frequencies, even when these are small, and assume that population sizes are infinite.

Let the frequencies of chromosomes AB, Ab, aB and ab be given by x_1, x_2, x_3 and x_4 and let w_{ij} be the fitness of a zygote consisting of chromosome types i and j. The derived scheme of genotype fitnesses and frequencies is shown in Table 1 in two forms. Table 1a lists them according to the gametic contributions from the parents while Table 1b lists them according to their states at the two loci. BODMER and PARSONS (1962), and others, showed that the equations for change in the x_i are given by

(1a) $$x'_1 = [x_1 w_1 - r(w_{14} x_1 x_4 - w_{23} x_2 x_3)]/\bar{w}$$
(1b) $$x'_2 = [x_2 w_2 + r(w_{14} x_1 x_4 - w_{23} x_2 x_3)]/\bar{w}$$
(1c) $$x'_3 = [x_3 w_3 + r(w_{14} x_1 x_4 - w_{23} x_2 x_3)]/\bar{w}$$
(1d) $$x'_4 = [x_4 w_4 - r(w_{14} x_1 x_4 - w_{23} x_2 x_3)]/\bar{w}$$

where r is the recombination fraction between the loci A and B,

(2) $$w_i = \sum_{j=1}^{4} x_j w_{ij}, \qquad \text{for } i = 1, 2, 3, 4$$

and

$$\bar{w} = x_1 w_1 + x_2 w_2 + x_3 w_3 + x_4 w_4$$
$$= \sum_{i=1}^{4} \sum_{j=1}^{4} x_i x_j w_{ij}.$$

TABLE 1

Fitnesses and frequencies of genotypes

(a)

	AB	Ab	aB	ab
AB	w_{11}	w_{12}	w_{13}	w_{14}
	$x_1\,x_1$	$x_1\,x_2$	$x_1\,x_3$	$x_1\,x_4$
Ab	w_{12}	w_{22}	w_{23}	w_{24}
	$x_2\,x_1$	$x_2\,x_2$	$x_2\,x_3$	$x_2\,x_4$
aB	w_{13}	w_{23}	w_{33}	w_{34}
	$x_3\,x_1$	$x_3\,x_2$	$x_3\,x_3$	$x_3\,x_4$
ab	w_{14}	w_{24}	w_{34}	w_{44}
	$x_4\,x_1$	$x_4\,x_2$	$x_4\,x_3$	$x_4\,x_4$

(b)

	BB	Bb	bb
AA	w_{11}	w_{12}	w_{22}
	x^2_1	$2x_1\,x_2$	x^2_2
Aa	w_{13}	w_{14}, w_{23}	w_{24}
	$2x_1\,x_3$	$2x_1\,x_4, 2x_2\,x_3$	$2x_2\,x_4$
aa	w_{33}	w_{34}	w_{44}
	x^2_3	$2x_3\,x_4$	x^2_4

The frequencies of the gametes AB, Ab, aB and ab are x_1, x_2, x_3 and x_4 respectively. w_{ij} is the fitness of the genotype formed by the combination of the ith with the jth gamete. Assuming random mating, this has a frequency $2x_i\,x_j$, $i \neq j$ and x^2_i, $i=j$.

These equations are similar to those derived by LEWONTIN and KOJIMA (1960) but allow for a difference between the fitnesses of coupling and repulsion double-heterozygotes. Note that the numbering of the gamete types differs from that in BODMER and PARSONS (1962).

We shall make use of the parameter $D = x_1\,x_4 - x_2\,x_3$, which is a measure of gametic interaction. The gamete frequencies can be expressed as functions of $p_1 = x_1 + x_2$, the frequency of A, $p_2 = x_1 + x_3$, the frequency of B, and D, the gametic interaction as follows:

$$x_1 = p_1\,p_2 + D$$
$$x_2 = p_1\,(1-p_2) - D$$
(3)
$$x_3 = (1-p_1)\,p_2 - D$$
$$x_4 = (1-p_1)\,(1-p_2) + D$$

D thus measures the difference between the frequency of the gamete AB and the frequency which would be obtained if alleles A and B were randomly associated in the population, and similarly for other gametic types. D has often been referred to as the "linkage disequilibrium" parameter, and more recently as the "gametic phase unbalance" (JAIN and ALLARD 1966).

2. *The Condition for the Existence of a Nontrivial Equilibrium with* $D = 0$

When $D = 0$, the gamete frequencies can be expressed as the products of the appropriate gene frequencies, so that the two loci are effectively independent. It

is a classical result due to JENNINGS (1917) that, in the absence of selection, the gamete frequencies approach this state at a rate $1 - r$. However in the presence of selection, even when $r = \frac{1}{2}$, this is no longer true as was pointed out by WRIGHT (1952) for a special set of symmetric viabilities. $|D|$ has a maximum value of $\frac{1}{4}$, when either $x_1 = x_4 = 0$ and $x_2 = x_3 = \frac{1}{2}$ or $x_1 = x_4 = \frac{1}{2}$ and $x_2 = x_3 = 0$, representing the maximum possible disturbances from independence between the loci. BODMER and PARSONS (1962) outlined the derivation of a general condition that must be satisfied by the selective values in order for a nontrivial equilibrium with $D = 0$ to exist (see also MORAN 1964 and LEWONTIN 1964a,b).

Assume for simplicity that $w_{14} = w_{23}$ so that equations (1) take the form

$$(4) \qquad \begin{aligned} \bar{w}\, x'_i &= x_i\, w_i - r\, w_{14}\, D, &&\text{for } i = 1, 4 \\ \bar{w}\, x'_i &= x_i\, w_i + r\, w_{14}\, D, &&\text{for } i = 2, 3 \end{aligned}$$

Then a nontrivial equilibrium with $D = 0$ will only exist if $w_i = \bar{w}$ for all i. These equations together with $D = 0$ and $x_1 + x_2 + x_3 + x_4 = 1$ impose a single condition on the viabilities which, using equations (3), is given by eliminating p_1, p_2 and $\bar{w}$ from the following four equations.

$$(5) \quad p_1 p_2\,(w_{i1} - w_{i2} - w_{i3} + w_{i4}) + p_1\,(w_{i2} - w_{i4}) + p_2\,(w_{i3} - w_{i4}) - \bar{w} = -w_{i4},$$
$$\text{for } i = 1, 2, 3, 4$$

Applying Cramer's rule for the solution of a set of simultaneous linear equations and expressing the fact that $p_1 p_2 = p_1 \times p_2$ the condition can be written in the form

$$(6) \quad \begin{vmatrix} w_{11}-w_{12}-w_{13}+w_{14} & w_{12}-w_{14} & w_{13}-w_{14} & -1 \\ w_{21}-w_{22}-w_{23}+w_{24} & w_{22}-w_{24} & w_{23}-w_{24} & -1 \\ w_{31}-w_{32}-w_{33}+w_{34} & w_{32}-w_{34} & w_{33}-w_{34} & -1 \\ w_{41}-w_{42}-w_{43}+w_{44} & w_{42}-w_{44} & w_{43}-w_{44} & -1 \end{vmatrix} \begin{vmatrix} w_{12}-w_{14} & w_{13}-w_{14} & -1 & -w_{14} \\ w_{22}-w_{24} & w_{23}-w_{24} & -1 & -w_{24} \\ w_{32}-w_{34} & w_{33}-w_{34} & -1 & -w_{34} \\ w_{42}-w_{44} & w_{43}-w_{44} & -1 & -w_{44} \end{vmatrix}$$

$$- \begin{vmatrix} w_{11}-w_{12}-w_{13}+w_{14} & w_{13}-w_{14} & -1 & -w_{14} \\ w_{21}-w_{22}-w_{23}+w_{24} & w_{23}-w_{24} & -1 & -w_{24} \\ w_{31}-w_{32}-w_{33}+w_{34} & w_{33}-w_{34} & -1 & -w_{34} \\ w_{41}-w_{42}-w_{43}+w_{44} & w_{43}-w_{44} & -1 & -w_{44} \end{vmatrix} \begin{vmatrix} w_{11}-w_{12}-w_{13}+w_{14} & w_{12}-w_{14} & -1 & -w_{14} \\ w_{21}-w_{22}-w_{23}+w_{24} & w_{22}-w_{24} & -1 & -w_{24} \\ w_{31}-w_{32}-w_{33}+w_{34} & w_{32}-w_{34} & -1 & -w_{34} \\ w_{41}-w_{42}-w_{43}+w_{44} & w_{42}-w_{44} & -1 & -w_{44} \end{vmatrix}$$
$$= 0$$

There is no obvious general parametrization for the selective values which encompasses this single condition although there are two interesting special cases, the "additive" and the "multiplicative" models for which it is satisfied.

3. The Additive Viability Model

When $w_{i1} - w_{i2} - w_{i3} + w_{i4} = 0$ for all i equation (6) is clearly satisfied. In this case the fitnesses can be expressed in the form given in Table 2 assuming the coupling and repulsion heterozygote fitnesses are equal, i.e., $w_{14} = w_{23}$. It can then easily be shown that

$$p_1 = (a_2 - a_3)/(2a_2 - a_1 - a_3) \qquad \text{and} \qquad p_2 = (b_2 - b_3)/(2b_2 - b_1 - b_3)$$

corresponding to the relevant solutions for the two loci considered separately, is a unique solution. Thus equations (4) give, at equilibrium when $x'_1 = x_i$, and dividing each by x_i and taking the second and third from the first and fourth equations

TABLE 2

Fitness scheme for the additive model

	BB	Bb	bb
AA	$w_{11} = a_1 + b_1$	$w_{12} = a_1 + b_2$	$w_{22} = a_1 + b_3$
Aa	$w_{13} = a_2 + b_1$	$w_{14}(=w_{23}) = a_2 + b_2$	$w_{24} = a_2 + b_3$
aa	$w_{33} = a_3 + b_1$	$w_{34} = a_3 + b_2$	$w_{44} = a_3 + b_3$

$$(7) \qquad w_1 - w_2 - w_3 + w_4 = r w_{14} \, D \left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} \right) = 0 \, ,$$

when the viabilities are additive. Since $1/x_1 + 1/x_2 + 1/x_3 + 1/x_4$, r and w_{14} are intrinsically positive we must have $D = 0$. Given $D = 0$ it is easy to determine the stated equilibrium as the unique internal solution of equations (5) and so of the general equations (1) for this additive model (see also MORAN 1967). The solution is, of course, valid only if $a_2 > a_1, a_3$ and $b_2 > b_1, b_3$. It will be shown later that in this case there exists no stable point on the boundaries of the gamete simplex, so that this solution is almost always stable when it exists. Only for very special combinations of the viabilities can one construct models giving no unique stable equilibrium.

4. *Multiplicative Viability Model*

The fitness scheme for the multiplicative model is illustrated in Table 3. It can easily be shown by substitution into equations (5) that

$$p_1 = (\alpha_2 - \alpha_3)/(2\alpha_2 - \alpha_1 - \alpha_3) \qquad p_2 = (\beta_2 - \beta_3)/(2\beta_2 - \beta_1 - \beta_3)$$

is a solution which satisfies the condition $D = 0$ (see also MORAN 1967). As for the additive case, this corresponds to the equilibria obtained separately for each of the two loci. It will, however, be shown below that in this case other internal equilibria can exist which may be stable when the equilibrium with $D = 0$ is not stable. The condition for stability depends on the recombination fraction r.

5. *A General Necessary Condition for the Stability of a Nontrivial Equilibrium with* $D = 0$

A necessary condition for the stability of a nontrivial equilibrium with $D = 0$ can be obtained by examining the behavior of equations (4) in the neighborhood of $D = 0$. Substituting from equations (3) into equations (2) we obtain

TABLE 3

Fitness scheme for the multiplicative model

	BB	Bb	bb
AA	$w_{11} = \alpha_1 \beta_1$	$w_{12} = \alpha_1 \beta_2$	$w_{22} = \alpha_1 \beta_3$
Aa	$w_{13} = \alpha_2 \beta_1$	$w_{14} = \alpha_2 \beta_2 = w_{23}$	$w_{24} = \alpha_3 \beta_3$
aa	$w_{33} = \alpha_3 \beta_1$	$w_{34} = \alpha_3 \beta_2$	$w_{44} = \alpha_3 \beta_3$

(8) $w_i = w_{i1}(p_1 p_2 + D) + w_{i2}[p_1(1-p_2) - D] + w_{i3}[(1-p_1)p_2 - D]$
$$+ w_{i4}[(1-p_1)(1-p_2) + D]$$

or

$$w_i = w_i{}^* + D\varepsilon_i, \qquad\qquad\qquad i = 1, 2, 3, 4$$

where

$$w_i{}^* = w_{i1}\, p_1\, p_2 + w_{i2}\, p_1(1-p_2) + w_{i3}(1-p_1)\, p_2 + w_{i4}(1-p_1)\,(1-p_2)$$

is the marginal fitness of the ith gamete when $D = 0$ and $\varepsilon_i = w_{i1} - w_{i2} - w_{i3} + w_{i4}$, is a measure of the additive viability interaction, which is zero for all i in the additive model. Thus

(9) $\displaystyle \bar{w} = \sum_{i=1}^{4} w_i\, x_i = (w_1{}^* + D\varepsilon_1)(p_1 p_2 + D) + (w_2{}^* + D\varepsilon_2)(p_1(1-p_2) - D)$
$$+ (w_3{}^* + D\varepsilon_3)((1-p_1)p_2 - D) + (w_4{}^* + D\varepsilon_4)((1-p_1)(1-p_2) + D)$$
$$= \bar{w}^* + 2D\bar{\varepsilon} + D^2(\varepsilon_1 - \varepsilon_2 - \varepsilon_3 + \varepsilon_4)$$

where

$$\bar{w}^* = w_1{}^*\, p_1 p_2 + w_2{}^*\, p_1(1-p_2) + w_3{}^*(1-p_1)p_2 + w_4{}^*(1-p_1)(1-p_2)$$

in the mean fitness when $D = 0$ and

$$\bar{\varepsilon} = \varepsilon_1 p_1 p_2 + \varepsilon_2 p_1(1-p_2) + \varepsilon_3(1-p_1)p_2 + \varepsilon_4(1-p_1)(1-p_2) = w_1{}^* - w_2{}^* - w_3{}^* + w_4{}^*.$$

Note that for the additive model, when $\varepsilon_i = 0$ for all i, $\bar{w} = \bar{w}^*$ for all values of p_1, p_2 and D. From equation (4) we have

(10) $\bar{w}^2\, D' = \bar{w}^2(x_1' x_4' - x_2' x_3')$
$$= (w_1 x_1 - r w_{14} D)(w_4 x_4 - r w_{14} D) - (w_2 x_2 + r w_{14} D)(w_3 x_3 + r w_{14} D)$$
$$= w_1 w_4 x_1 x_4 - w_2 w_3 x_2 x_3 - r w_{14}\, \bar{w}\, D.$$

Substituting into this equation from equation (8) we obtain

(11) $\bar{w}^2\, D' = p_1(1-p_1)\, p_2(1-p_2)(w_1{}^* w_4{}^* - w_2{}^* w_3{}^*)$
$$+ D[p_1(1-p_1)\, p_2(1-p_2)(\varepsilon_4 w_1{}^* + \varepsilon_1 w_4{}^* - \varepsilon_3 w_2{}^* - \varepsilon_2 w_3{}^*)$$
$$+ [p_1(1-p_2) + p_2(1-p_1)]\,(w_2{}^* w_3{}^* - w_1{}^* w_4{}^*)$$
$$+ w_1{}^* w_4{}^* - r w_{14}\, \bar{w}]$$
$$+ D^2[p_1(1-p_1)\, p_2(1-p_2)(\varepsilon_1 \varepsilon_4 - \varepsilon_2 \varepsilon_3) + w_1{}^* w_4{}^* - w_2{}^* w_3{}^*$$
$$+ \varepsilon_1 w_4{}^* + \varepsilon_4 w_1{}^* + [p_1(1-p_2) + p_2(1-p_1)]\,[\varepsilon_2 w_3{}^*$$
$$+ \varepsilon_3 w_2{}^* - \varepsilon_1 w_4{}^* - \varepsilon_4 w_1{}^*]]$$
$$+ D^3[(p_1(1-p_2) + p_2(1-p_1))(\varepsilon_2 \varepsilon_3 - \varepsilon_1 \varepsilon_4) + \varepsilon_1 \varepsilon_4 + \varepsilon_1 w_4{}^*$$
$$+ \varepsilon_4 w_1{}^* - \varepsilon_2 w_3{}^* - \varepsilon_3 w_2{}^*]$$
$$+ D^4(\varepsilon_1 \varepsilon_4 - \varepsilon_2 \varepsilon_3).$$

KIMURA (1965) has indicated by an analysis of an equation analogous to equation (11) together with some numerical simulations that when r is near 0.5 or fitness interactions are small, a state near $D = 0$ is rapidly approached and FISHER's fundamental theorem holds to a good degree of approximation.

At a nontrivial equilibrium with $D = 0$, from equations (4) and (8) $w_1{}^* = w_4{}^* = w_2{}^* = w_3{}^* = \bar{w}$. Thus if D is perturbed from an equilibrium with $D = 0$ keeping p_1 and p_2 constant at their equilibrium values, it will change in the following generation to a value given by

(12) $\bar{w}^2\, D' = D[p_1(1-p_1)\, p_2(1-p_2)\, \bar{w}(\varepsilon_1 - \varepsilon_2 - \varepsilon_3 + \varepsilon_4) + \bar{w}^2 - r w_{14}\, \bar{w}] + 0(D^2)$

where $\bar{w}, p_1$ and p_2 take their equilibrium values. A necessary condition for the equilibrium to be stable is, therefore, that $|D'| < |D|$. When the coefficient of D on the right hand side of equation (12) is positive, we have

$$\bar{w}^2 > p_1(1-p_2)\, p_2(1-p_2)\bar{w}(\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4) + \bar{w}^2-rw_{14}\,\bar{w}$$

which gives

(13) $$r > p_1(1-p_1)\, p_2(1-p_2)\, (\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4)/w_{14}\,.$$

If, on the other hand,

$$p_1(1-p_1)\, p_2(1-p_2)\, \bar{w}(\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4) + \bar{w}^2 - rw_{14}\,\bar{w} < 0$$

then

$$r > [p_1(1-p_1)\, p_2(1-p_2)\, (\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4) + \bar{w}]/w_{14}$$
$$> p_1(1-p_1)\, p_2(1-p_2)\, (\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4)/w_{14}$$

since $\bar{w}$ and w_{14} are intrinsically positive. The condition $|D'/D| < 1$ now takes the form

$$\bar{w}^2 > rw_{14}\,\bar{w} - \bar{w}^2 - p_1(1-p_1)\, p_2(1-p_2)\, \bar{w}(\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4)$$

which reduces to

$$r < \frac{p_1(1-p_1)\, p_2(1-p_2)\, (\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4)}{w_{14}} + 2\,\frac{\bar{w}}{w_{14}}\,.$$

The equilibrium with $D = 0$ can, therefore, only be stable if r lies between the values

(13a)
$$\frac{p_1(1-p_1)p_2(1-p_2)\,(\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4)}{w_{14}}\text{ and }\frac{p_1(1-p_1)p_2(1-p_2)\,(\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4)}{w_{14}}+2\frac{\bar{w}}{w_{14}}.$$

Since $\bar{w}/w_{14}$ will, for most reasonable sets of fitness values, not deviate markedly from 1, condition (13) will generally predominate. So long as $\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4 > 0$, there will always exist a lower limit on r below which the linkage equilibrium cannot be stable. For the additive model, $\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4 = 0$ and condition (13) is always satisfied.

It can easily be shown that for the multiplicative model
$$w_1{}^*w_4{}^* - w_2{}^* \, w_3{}^* = 0$$
for all values of p_1 and p_2 so that, from equation (11), $D = 0$ is an eigenvector for linear perturbations of D, p_1, and p_2 from their equilibrium values. Thus (13a) gives the condition for D to increase from $D = 0$ in all directions, that is for all values of p_1 and p_2. Since the stabilities of the gene frequencies to perturbations from their equilibrium values when $D = 0$ are guaranteed by the heterotic nature of the fitness scheme, required for this equilibrium to exist, the eigenvalues for the eigenvectors orthogonal to $D = 0$ must both be less than unity, so that condition (13a) now becomes a necessary and sufficient condition for stability of the equilibrium with $D = 0$. The fitnesses given in Table 3 can be written, without loss of generality, in the form $\alpha_2 = \beta_2 = 1$, $\alpha_1 = 1-s_1$, $\alpha_3 = 1-t_1$, $\beta_1 = 1-s_2$, $\beta_3 = 1-t_2$ so that the equilibrium is at

$$p_1 = t_1/s_1+t_1 \qquad p_2 = t_2/s_2+t_2.$$

We then have

$$\varepsilon_1 = w_{11} - w_{12} - w_{13} + w_{14} = \alpha_1\beta_1 -\alpha_1\beta_2 - \alpha_2\beta_1 + \alpha_2\beta_2$$
$$= (1-s_1)\, (1-s_2) -1 + s_1-1 + s_2 + 1 = s_1s_2$$

and similarly for ε_2, ε_3 and ε_4 so that

$$\varepsilon_1 - \varepsilon_2 - \varepsilon_3 + \varepsilon_4 = (s_1+t_1)\, (s_2+t_2)$$

and condition (13) takes the form

(14) $$r > \left(\frac{s_1 t_1}{s_1 + t_1} \right) \left(\frac{s_2 t_2}{s_2 + t_2} \right).$$

Now $s_1 t_1/(s_1+t_1)$ and $s_2 t_2/(s_2+t_2)$ are the "marginal" segregation loads at loci A and B respectively. Thus, two multiplicative overdominant loci cannot be at a stable nontrivial equilibrium with $D = 0$ if the recombination fraction between them is less than the product of their segregational loads.

LEWONTIN's (1964b) second five-locus model involves multiplicative determination of fitnesses. It is interesting to note that, although he finds that the equilibrium involving $D = 0$ becomes unstable at about $r = 0.065$, the above formula (14) predicts that it should become unstable at $r = 0.0625$. This shows that with more than two loci segregating, higher-order gene interactions play some part, although perhaps only a small one, in further increasing the minimum value of r necessary for the stability of an equilibrium with $D = 0$.

6. A General Symmetrical Viability Model

6.1 *Derivation of the equilibrium solutions.* It has not, so far, been possible to obtain general explicit solutions for the equilibria given by the equations (1). A number of authors (WRIGHT 1952; KIMURA (1956; LEWONTIN and KOJIMA 1960; BODMER and PARSONS 1962) have however considered simplified symmetric viability models which can be completely solved. A symmetrical fitness scheme which includes essentially all of these models is given in Table 4. When $\alpha = \delta$ it corresponds to LEWONTIN and KOJIMA's (1960) model, when $\beta = \gamma$ it corresponds to BODMER and PARSON's (1962) model (of which WRIGHT's (1952) is a special case) and when $\delta = -s+t$, $\alpha = s+t$, $\beta = 0$ and $\gamma = t$ it corresponds to KIMURA's (1956) model.

For this model equation (1) takes the form

(15a) $\bar{w}_1' x_1' = x_1 - \delta x_1{}^2 - \beta x_1 x_2 - \gamma x_1 x_3 - rD$

(15b) $\bar{w}_2' x_2' = x_2 - \beta x_1 x_2 - \alpha x_2{}^2 - \gamma x_2 x_4 - rD$

(15c) $\bar{w} x_3' = x_3 - \gamma x_1 x_3 - \alpha x_3{}^2 - \beta x_3 x_4 + rD$

(15d) $\bar{w} x_4' = x_4 - \gamma x_2 x_4 - \beta x_3 x_4 - \delta x_4{}^2 - rD$

where

$\bar{w} = 1 - \delta(x_1{}^2 + x_4{}^2) - \alpha(x_2{}^2 + x_3{}^2) - 2\beta(x_3 x_4 + x_1 x_2) - 2\gamma(x_1 x_3 + x_2 x_4).$

At equilibrium when $x_i' = x_i$, subtracting (15a) from (15d) and (15b) from (15c) gives

(16) $\bar{w}(x_1 - x_4) = x_1 - x_4 - \delta(x_1{}^2 - x_4{}^2) - \beta(x_1 x_2 - x_3 x_4) - \gamma(x_1 x_3 - x_2 x_4)$

$\bar{w}(x_2 - x_3) = x_2 - x_3 - \alpha(x_2{}^2 - x_3{}^2) - \beta(x_1 x_2 - x_3 x_4) + \gamma(x_1 x_3 - x_2 x_4)$

TABLE 4

Fitness scheme for general symmetrical viability model

	BB	Bb	bb
AA	$w_{11} = 1-\delta$	$w_{12} = 1-\beta$	$w_{22} = 1-\alpha$
Aa	$w_{13} = 1-\gamma$	$w_{14} = 1 = w_{23}$	$w_{24} = 1-\gamma$
aa	$w_{33} = 1-\alpha$	$w_{34} = 1-\beta$	$w_{44} = 1-\delta$

so that $x_1 = x_4$, $x_2 = x_3$ clearly gives an equilibrium solution of equations (15). It can be shown that when $\beta = \gamma > 0$ and $\alpha, \delta > 0$ no other equilibrium can exist in which $x_1 \neq x_4$ and $x_2 \neq x_3$. However, equilibria with $x_1 = x_4$ and $x_2 \neq x_3$ (or vice versa) do exist for certain special values of the fitnesses. When α, β, γ, δ are not all positive, equilibria in which $x_1 \neq x_4$ and $x_2 \neq x_3$ also, sometimes, exist. Details of these analyses will be published elsewhere (BODMER, M. FELDMAN and S. KARLIN, in preparation). It seems likely, however, that in general equilibria in which either $x_1 \neq x_4$ or $x_2 \neq x_3$ are of little biological significance since their existence depends on special relationships holding between the viabilities.

When $x_1 = x_4$ and $x_2 = x_3$ we have

$$x_1 + x_2 = x_3 + x_4 = \tfrac{1}{2} = p_1 = 1 - p_1$$

and

$$x_1 + x_3 = x_2 + x_4 = \tfrac{1}{2} = p_2 = (1 - p_2).$$

Substituting $x_1 = \tfrac{1}{4} + D = x_4$ and $x_2 = \tfrac{1}{4} - D = x_3$ in equation (15a) gives the following cubic equation for the value of D at equilibrium,

$$(17) \qquad 64\, lD^3 - 16mD^2 - 4(l - 8r)\, D + m = 0$$

where $l = 2(\beta + \gamma) - (\alpha + \delta)$ and $m = (\delta - \alpha)$. This equation is equivalent to equation (7) of BODMER and PARSONS (1962), since β and γ only occur in the form $\beta + \gamma$ so that setting $\beta = \gamma$ results in no loss of generality for the model as far as the equilibrium solutions with $x_1 = x_4$ and $x_2 = x_3$ are concerned. A number of special cases of equation (17) are easily solved and will be considered next.

When $r = 0$, so that the system is equivalent to a four allele model, equation (17) reduces to

$$(18) \qquad (16\, D^2 - 1)\, (4lD - m) = 0$$

giving $D = +\tfrac{1}{4}, -\tfrac{1}{4}$ and $m/4l$.

The first two solutions correspond respectively to the equilibria $x_2 = x_3 = 0$ and $x_1 = x_4 = 0$, when either gametes aB and ab or AB and Ab are absent from the population.

If $m = 0$ or $\delta = \alpha$, which is the case considered by LEWONTIN and KOJIMA (1960), equation (17) reduces to

$$(19) \qquad D(64\, lD^2 - 4(l - 8r)) = 0$$

giving $D = 0$ (i.e., $x_1 = x_2 = x_3 = x_4 = \tfrac{1}{4}$) or $D = \pm \tfrac{1}{4}(1 - (8r/l))^{\frac{1}{2}}$. The solutions other than $D = 0$ are only valid if $|D| < \tfrac{1}{4}$ that is, if $l > 0$. $D = 0$ is only a root of equation (17) if $m = 0$, which in this case is therefore a necessary condition for the existence of "linkage equilibrium."

The case considered by WRIGHT (1952) corresponds to $l = 0$, when equation (17) reduces to

$$(20) \qquad 16\, mD^2 - 32\, rD - m = 0$$

giving solutions

$$D = \frac{r}{m} \pm \left(\frac{16\, r^2 + m^2}{16\, m^2}\right)^{\frac{1}{2}} = \frac{r}{m} + \tfrac{1}{4}\left(1 + 16\,\frac{r^2}{m^2}\right)^{\frac{1}{2}}.$$

When r/m is small, we have $D = \pm\tfrac{1}{4} + r/m + 0(r^2/m^2)$. On the other hand, when r/m is large (m/r small) $D = (1/32)(m/r) + 0(m^2/r^2)$.

Following BODMER and PARSONS (1962), useful approximate solutions to equation (17) can also be obtained when r is small and $l \neq 0$, $m \neq 0$. Differenti-

ating equation (17) implicitly with respect to r gives

$$192\ lD^2\ \frac{dD}{dr} - 32\ mD\ \frac{dD}{dr} - 4(l-8r)\ \frac{dD}{dr} + 32\ D = 0$$

or

(21) $$dD/dr = 8D/(l - 8r + 8mD - 48lD^2).$$

Using Taylor's theorem in the neighborhood of the three solutions of equations (18), namely, substituting $r = 0$ and $D = \frac{1}{4}$, $-\frac{1}{4}$ and $m/4l$ into equation (21) gives the following approximate solutions, to $0(r^2)$,

(22a) $\quad D = \frac{1}{4} - \dfrac{r}{l-m}$; $\qquad x_1 = x_4 = \frac{1}{2} - \dfrac{r}{l-m}$, $\qquad m \neq l$

(22b) $\quad D = -\frac{1}{4} + \dfrac{r}{l+m}$; $\qquad x_1 = x_4 = \dfrac{r}{m+l}$, $\qquad m \neq -l$

(22c) $\quad D = \dfrac{m}{4l} + \dfrac{2m}{l^2-m^2}$; $\qquad x_1 = x_4 = \frac{1}{4} + \dfrac{m}{4l} + \dfrac{2m}{l^2-m^2}$

Note that solution (22a) is only valid when $l-m > 0$ or $\beta + \gamma > \delta$ and (22b) is only valid if $l+m > 0$ or $\beta + \gamma > \alpha$. The first and second solutions correspond to equilibria in which either gametes AB and ab or Ab and aB are in the majority, and represent extreme cases of linkage "disequilibrium". The conditions for their existence correspond to the type of optimum model proposed by MATHER (1941 and later), as discussed by BODMER and PARSON (1962). The third solution corresponds, in general, to an equilibrium near $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$, or near $D = 0$, provided m/l is small. Numerical examples of solutions of equation (17) have been given by WRIGHT (1952), KIMURA (1956), LEWONTIN and KOJIMA (1960), BODMER and PARSONS (1962), PARSONS (1963a), and JAIN and ALLARD (1966).

6.2. *Conditions for the stability of the equilibria.* In order to investigate the stability of an equilibrium given by $x_1 = x_4 = x$, $x_2 = x_3 = \frac{1}{2} - x$ (i.e., $D = -\frac{1}{4} + x$), we examine the behavior of equations (15) in the neighborhood of the equilibrium. Substituting into these equations

$$x_1 = x + X_1,\quad x_2 = \frac{1}{2} - x - X_2,\quad x_3 = \frac{1}{2} - x - X_3,\quad x_4 = x + X_4$$

where $X_1 + X_4 = X_2 + X_3$, and ignoring quadratic terms in X_1, X_2, X_3 and X_4 gives the following set of linear equations relating X'_i to X_i, $i = 1,2,3,4$:

(23a) $\quad \bar{w}_e X'_1 = X_1[1 - 2\delta x - (\frac{1}{2} - x)(\beta + \gamma) + 2\delta x^2 + 2x(\frac{1}{2} - x)(\beta + \gamma)]$
$\qquad\qquad + X_2[\beta x - r/2 - 2x(\frac{1}{2} - x)\alpha - 2x^2(\beta + \gamma)]$
$\qquad\qquad + X_2[\gamma x - r/2 - 2x(\frac{1}{2} - x)\alpha - 2x^2(\beta + \gamma)]$
$\qquad\qquad + X_4[2\delta x^2 + 2x(\frac{1}{2} - x)(\beta + \gamma)]$

(23b) $\quad \bar{w}_e X'_2 = X_1[\beta(\frac{1}{2} - x) - 2\delta x(\frac{1}{2} - x) - 2(\frac{1}{2} - x)^2(\beta + \gamma)]$
$\qquad\qquad + X_2[1 - (\beta + \gamma)x - 2\alpha(\frac{1}{2} - x) - r/2 + 2\alpha(\frac{1}{2} - x)^2$
$\qquad\qquad\qquad + 2x(\frac{1}{2} - x)(\beta + \gamma)]$
$\qquad\qquad + X_3[-r/2 + 2(\frac{1}{2} - x)^2\alpha + 2x(\frac{1}{2} - x)(\beta + \gamma)]$
$\qquad\qquad + X_4[\gamma(\frac{1}{2} - x) - 2\delta x(\frac{1}{2} - x) - 2(\frac{1}{2} - x)^2(\beta + \gamma)]$

(23c) $\quad \bar{w}_e X'_3 = X_1[\gamma(\frac{1}{2} - x) - 2\delta x(\frac{1}{2} - x) - 2(\frac{1}{2} - x)^2(\beta + \gamma)]$
$\qquad\qquad + X_2[-r/2 + 2(\frac{1}{2} - x)^2\alpha + 2x(\frac{1}{2} - x)(\beta + \gamma)]$
$\qquad\qquad + X_3[1 - (\beta + \gamma)x - 2\alpha(\frac{1}{2} - x) - r/2 + 2\alpha(\frac{1}{2} - x)^2$
$\qquad\qquad\qquad + 2x(\frac{1}{2} - x)(\beta + \gamma)]$

$$+X_4[\beta(\tfrac{1}{2}-x)-2x(\tfrac{1}{2}-x)\delta-2(\tfrac{1}{2}-x)^2(\beta+\gamma)]$$

(23d) $\quad \bar{w}_e X'_4 = X_1[2\delta x^2 + 2x(\tfrac{1}{2}-x)(\beta+\gamma)]$

$$+X_2[\gamma x - (r/2) - 2x(\tfrac{1}{2}-x)\alpha - 2x^2(\beta+\gamma)]$$

$$+X_3[\beta x - (r/2) - 2x(\tfrac{1}{2}-x)\alpha - 2x^2(\beta+\gamma)]$$

$$+X_4[1 - 2\delta x - (\tfrac{1}{2}-x)(\beta+\gamma) + 2\delta x^2 + 2x(\tfrac{1}{2}-x)(\beta+\gamma)]$$

where $\bar{w}_e = 1 - 2\delta x^2 - 2\alpha(\tfrac{1}{2}-x)^2 - 4(\beta+\gamma)x(\tfrac{1}{2}-x)$ is the value of $\bar{w}$ at equilibrium.

Adding and subtracting equations (23a) and (23d), and (23b) and (23c) gives

(24a) $\quad \bar{w}_e(X'_1 + X'_4) = (X_1 + X_4)(1 - 2\delta x - (\tfrac{1}{2}-x)(\beta+\gamma) + 4\delta x^2$
$+ 4x(\tfrac{1}{2}-x)(\beta+\gamma)) + (X_2 + X_3)((\beta+\gamma)x - r - 4x(\tfrac{1}{2}-x)\alpha - 4x^2(\beta+\gamma))$

(24b) $\quad \bar{w}_e(X'_2 + X'_3) = (X_1 + X_4)((\beta+\gamma)(\tfrac{1}{2}-x) - 4x(\tfrac{1}{2}-x)\delta$
$$-4(\tfrac{1}{2}-x)^2(\beta+\gamma))$$
$$+ (X_2 + X_3)(1 - (\beta+\gamma)x - 2\alpha(\tfrac{1}{2}-x) + 4\alpha(\tfrac{1}{2}-x)^2$$
$$+ 4x(\tfrac{1}{2}-x)(\beta+\gamma))$$

(24c) $\quad \bar{w}_e(X'_1 - X'_4) = (X_1 - X_4)(1 - 2\delta x - (\tfrac{1}{2}-x)(\beta+\gamma))$
$$+ (X_2 - X_3)(\beta-\gamma)x$$

(24d) $\quad \bar{w}_e(X'_2 - X'_3) = (X_1 - X_4)(\tfrac{1}{2}-x)(\beta-\gamma)$
$$+ (X_2 - X_3)(1 - 2\alpha(\tfrac{1}{2}-x) - x(\beta+\gamma))$$

It is easily seen that equations (24a) and (24b) are equivalent since $X_1 + X_2 = X_3 + X_4$. The system of linear equations (23) can thus be transformed into equations (24c) and (24d), involving only $X_1 - X_4$ and $X_2 - X_3$ and the following equation in $Z = X_1 + X_2 = X_3 + X_4$,

(25) $\qquad \bar{w}_e Z' = Z(1 - r + 2x(1-2x)l - \tfrac{1}{2}(\beta+\gamma)).$

The equilibrium will, therefore, be stable provided $|Z'/Z| < 1$ and the characteristic roots of the pair of linear equations (24c) and (24d) both have modulus less than unity. When the coefficient of Z on the right hand side of (25) is positive, the condition for Z gives

(26) $\qquad \bar{w}_e > 1 - r + 2x(1-2x)l - \tfrac{1}{2}(\beta+\gamma)$ or
$$r > 1 + 2x(1-2x) - \tfrac{1}{2}(\beta+\gamma) - \bar{w}_e.$$

When, on the other hand,
$$1 - r + 2x(1-2x)l - \tfrac{1}{2}(\beta+\gamma) < 0,$$
the condition $|Z'/Z| < 1$ gives
$$\bar{w}_e > r - 1 - 2x(1-2x)l + \tfrac{1}{2}(\beta+\gamma),$$
since $\bar{w}_e$ is positive. These two conditions combine to give

(26a) $\quad 1 + 2x(1-2x)l - \tfrac{1}{2}(\beta+\gamma) < r < 1 + 2x(1-2x)l - \tfrac{1}{2}(\beta+\gamma) + \bar{w}_e.$

Since, in general, $\bar{w}_e$ will not be much less than one, and $\beta, \gamma < 1$, the upper limit on r in (26a) is of little significance and thus condition (26) generally predominates. Substituting for $\bar{w}_e$ into condition (26) gives

(27) $\qquad r > - (l+m)/4 + (3l+m)x - 6lx^2$

where l and m are defined as before. Thus, since equations (24c) and (24d) do not involve the recombination fraction r, (27) is, essentially, the only condition which relates the value of r to the stability of the equilibrium, but it is by itself only a necessary and not a sufficient condition for stability of the equilibrium. This condition is equivalent to equation (8) of BODMER and PARSONS (1962). The

characteristic equation for equations (24c) and (24d) reduces after some manipulation, to

$$(28) \quad \lambda^2 \bar{w}_e{}^2 - \lambda \bar{w}_e [2 - 2\delta x - 2\alpha(\tfrac{1}{2} - x) - \tfrac{1}{2}(\beta + \gamma)] + 1 + 4x(\tfrac{1}{2} - x)\beta\gamma$$
$$+ 4\delta\alpha x(\tfrac{1}{2} - x) - 2\delta x - 2\alpha(\tfrac{1}{2} - x) - \tfrac{1}{2}(\beta + \gamma)$$
$$+ 2(\beta + \gamma)(\delta x^2 + \alpha(\tfrac{1}{2} - x)^2) = 0,$$

which can be used to check the stability of an equilibrium in conjunction with condition (27). Equation (28) has, in general, no simple rational solution. When, however, $\alpha = \delta$ and $x = \tfrac{1}{4}$ $(D = 0)$, corresponding to one of the equilibria discussed by Lewontin and Kojima (1960), it has the solutions

$$(1/\bar{w}_e)(1 - (\beta + \alpha)/2) \qquad \text{and} \qquad (1/\bar{w}_e)(1 - (\gamma + \alpha)/2)$$

yielding the two stability conditions

$$\alpha > \gamma - \beta \qquad \text{and} \qquad \alpha > \beta - \gamma, \text{ since } \beta, \gamma, \alpha < 1,$$

which are equivalent to

$$(29) \qquad\qquad \alpha > 0 \qquad \text{and} \qquad \alpha > |\beta - \gamma|.$$

Substituting $\alpha = \delta$ and $x = \tfrac{1}{4}$ into condition (27) gives $r > l/8$, which together with (29) gives the complete sufficient condition for stability of this equilibrium (Lewontin and Kojima 1960). It is interesting to note that $r \le l/8$ is the condition for the existence of the other two solutions of equation (19).

The multiplicative model given in Table 3 reduces to a special case of the symmetrical model with $m = \delta - \alpha = 0$ when $\alpha_2 = \beta_2 = 1$ and $\alpha_1 = \alpha_3 = 1 - s_1$, $\beta_1 = \beta_3 = 1 - s_2$. The conditions for the stability of the equilibrium with $D = 0$, namely (29) together with $r > l/8$, reduce to

$$(30) \qquad r > \tfrac{1}{4} s_1 s_2, \quad (1 - s_1)(1 - s_2) < 1 \quad \text{and} \quad s_1 + s_2 - s_1 s_2 > |s_1 - s_2|$$

so that it is clear that, for the multiplicative viability model, if r is sufficiently small, stable nontrivial equilibria with $D = 0$ cannot exist, and so stable equilibria with $D \ne 0$ will usually exist. As pointed above, this is not true for the additive fitness model given in Table 2.

When r is small, substituting the approximate solutions (22a) and (22b) into condition (27) gives, to $0(r^2)$,

$$r < (l-m)^2/8l \qquad \text{and} \qquad r < (l+m)^2/8l$$

respectively, as necessary conditions for the stability of these equiilbria. When $r = 0$ and $x = 0$, equation (28) has solutions

$$\frac{1 - \alpha}{1 - \alpha/2} \quad \text{and} \quad \frac{1 - (\beta + \gamma)/2}{1 - \alpha/2}$$

both of which are less than one, if $\beta + \gamma > \alpha > 0$, which is the condition for the existence of the equilibrium given equation (22b). Thus provided r is small enough, stable equilibria in the neighborhood of $x = 0$ and $x = \tfrac{1}{2}$ will always exist so long as $\beta + \gamma > \alpha$ and $\beta + \gamma > \delta$, respectively.

7. Conditions for the Increase of a New Allele Linked to a Polymorphic Locus

We assume that, initially, gametes aB and ab are maintained polymorphic by an advantage of the heterozygote aB/ab over both homozygotes aB/aB and ab/ab, and examine the fate of gametes AB and Ab following their introduction into the population at low frequencies. We shall obtain the conditions for the increase in

frequency of gametes AB and ab by approximating to the general equations (1) in the neighborhood of the equilibrium given by $x_1 = x_2 = 0$. The equilibrium frequencies u and v of gametes aB and ab in the absence of AB and ab are given by

$$u = 1 - v = (w_{34} - w_{44})/(2w_{34} - w_{33} - w_{44})$$

where $w_{34} > w_{44}$ and w_{33} for the equilibrium to be stable before the introduction of A.

Following BODMER (BODMER and PARSONS 1962), we substitute

$$x_3 = u - d_3, \qquad x_4 = v - d_4, \qquad (x_1 + x_2 = d_3 + d_4)$$

into equations (1) and, assuming x_1, x_2, d_3, d_4 are small, ignore quadratic terms in x_1, etc. to obtain the following linear equations relating x'_1 etc. to x_1, etc.

(31a) $wx'_1 = x_1[w_1{}^* - rw_{14}v] + x_2 rw_{23}u$

(31b) $wx'_2 = x_1 rw_{14}v + x_2(w^*_2 - ruw_{23})$

(31c) $wd'_3 = x_1[uw_{13} + 2vw_{14} - rvw_{14}] + x_2[uw_{23} + 2vw_{24} + ruw_{23}]$
$\qquad\qquad + d_3[(1 - 2u)(uw_{33} + vw_{34}) + uw_{33}]$
$\qquad\qquad + d_4[uw_{34} - 2u[uw_{34} + vw_{44}]]$

(31d) $wd'_4 = x_1[2uw_{13} + vw_{14} + rvw_{14}] + x_2[2uw_{23} + vw_{24} - ruw_{23}]$
$\qquad\qquad + d_3[vw_{34} - 2v[uw_{33} + vw_{34}]] + d_4[vw_{44} +$
$\qquad\qquad (1 - 2u)(uw_{34} + vw_{34})]$

where

$$w = u^2 w_{33} + 2uvw_{34} + v^2 w_{44}$$

is the mean fitness before the introduction of the new gametes and

$$w_1{}^* = uw_{13} + vw_{14}, \qquad w_2{}^* = uw_{23} + vw_{24}$$

are the initial marginal fitnesses of the new gametes AB and Ab respectively.

Since equations (31a) and (31b) do not involve d_3 and d_4 the characteristic equation for equations (31) breaks down into two quadratics, the first corresponding to equations (31a) and (31b) and the second corresponding to the terms involving only d_3 and d_4 in equations (31c) and (31d). This latter quadratic is given by the determinental equation

(32) $$\begin{vmatrix} 2uv(w_{33} - w_{34}) + vw_{34} - w\lambda & 2uv(w_{34} - w_{44}) - uw_{34} \\ 2uv(w_{34} - w_{33}) - vw_{34} & 2uv(w_{44} - w_{34}) + uw_{34} - w\lambda \end{vmatrix} = 0$$

and has roots

$$\lambda_1 = 0 \quad \text{and} \quad \lambda_2 = (w_{34} - 2uv[w_{34} - w_{33} - w_{44}])/w.$$

If we write

$$s = 1 - w_{33}/w_{34} \quad \text{and} \quad t = 1 - w_{44}/w_{34},$$

following the usual notation for a two-allele balanced polymorphism, the second root becomes

$$\lambda_2 = \frac{1 - 2uv(s + t)}{1 - (st/s + t)} = \frac{1 - (2st/s + t)}{1 - (st/s + t)} = 1 - \frac{st}{s + t - st}.$$

Thus $\lambda_2 < 1$ provided s, $t > 0$ or $w_{34} > w_{33}$, w_{44}, which is the condition for the stability of the polymorphism in aB and ab in the absence of AB and Ab.

The remaining quadratic obtained from equations (31a) and 31b) is given by

$$\begin{vmatrix} w_1{}^* - rw_{14}v - \lambda & rw_{23}u \\ rw_{14}v & w_2{}^* - rw_{23}u - \lambda w \end{vmatrix} = 0$$

which reduces to

$$(33) \qquad w^2\lambda^2 - \lambda w[w_1{}^* + w_2{}^* - r(uw_{23} + vw_{14})] + w_1{}^*w_2{}^* \\ - r[uw_{23}w_1{}^* + vw_{14}w_2{}^*] = 0 .$$

The discriminant of this quadratic ("$b^2 - 4ac$") is given by

$$w^2[w_1{}^* + w_2{}^* - r(uw_{23} + vw_{14})]^2 - 4w^2[w_1{}^*w_2{}^* - r(uw_{23}w_1{}^* + vw_{14}w_2{}^*)] \\ = [w_1{}^* - w_2{}^* + r(uw_{23} - vw_{14})]^2 + 4r^2\,uvw_{23}w_{14} > 0 ,$$

since all quantities involved are intrinsically positive, so that the roots of equation (33) are always real. The condition for the increase of the new gametes is, therefore, that at least one of the two roots of this equation have modulus greater than unity. It is easily shown that for a quadratic equation given in the form $\lambda^2 + A\lambda + B = 0$, this condition is $1 + B - |A| < 0$ if $|A| < 2$, or $|A| > 2$. Since

$$r < 1 < (w_1{}^* + w_2{}^*)/(uw_{23} + vw_{14})$$

application of this condition to equation (33) gives

$$(34a) \qquad w_1{}^* + w_2{}^* - r(uw_{23} + vw_{14}) > 2w$$

or

$$(34b) \qquad (w - w_1{}^*)(w - w_2{}^*) + r[vw_{14}(w - w_2{}^*) - uw_{23}(w_1{}^* - w)] < 0$$

if

$$w_1{}^* + w_2{}^* - r(uw_{23} + vw_{14}) < 2w .$$

When $w_1{}^* = w_2{}^* = w^*$, say, equation (33) has roots w^*/w and

$$(w^* - r(uw_{23} + vw_{14}))/w ,$$

so that the dominant latent root is then always greater than one if $w^* > w$, which is therefore the condition for increase of the new gametes.

If $w > w_1{}^*$ and $w_2{}^*$, neither of conditions (34a) or (34b) can be satisfied, and the new gametes can never increase. If, on the other hand $w < w_1{}^*$, $w_2{}^*$ either (34a) is satisfied or

$$r > (w_1{}^* + w_2{}^* - 2w)/(uw_{23} + vw_{14}) .$$

Now

$$\frac{w_1{}^* - w + w_2{}^* - w}{uw_{23} + vw_{14}} - \frac{(w_1{}^* - w)(w_2{}^* - w)}{uw_{23}(w_1{}^* - w) + vw_{14}(w_2{}^* - w)}$$

$$= \frac{uw_{23}(w_1{}^* - w)^2 + vw_{14}(w_2{}^* - w)^2}{(uw_{23} + vw_{14})(uw_{23}(w_1{}^* - w) + vw_{14}(w_2{}^* - w))} > 0 ,$$

so that

$$r > (w_1{}^* + w_2{}^* - 2w)/(uw_{23} + vw_{14})$$

implies

$$r > \frac{(w_1{}^* - w)(w_2{}^* - w)}{(uw_{23}(w_1{}^* - w) + vw_{14}(w_2{}^* - w))}$$

which is condition (34b). Thus $w < w_1{}^*$, $w_2{}^*$ implies condition (34), and the new gametes always increase. Adding equations (31a) and (31b) gives

$$w(x'_1 + x'_2) = x_1 w_1{}^* + x_2 w_2{}^*,$$

from which it is easily seen that $p_1 = x_1 + x_2$, the gene frequency of A, always

increases if $w < w_1^*$, w_2^* and always decreases if $w > w_1^*$, w_2^*, verifying the above conditions. When $w_1^* > w > w_2^*$, the new gametes always increase if $vw_{14}(w-w_2^*) \leq uw_{23}(w_1^*-w)$, for then the left hand side of the first inequality of (34b) is always negative. If, however, $vw_{14}(w-w_2^*) > uw_{23}(w_1^*-w)$, the new gametes only increase if

$$r < \frac{(w-w_2^*)(w_1^*-w)}{vw_{14}(w-w_2^*)-uw_{23}(w_1^*-w)}.$$

A similar set of conditions holds when $w_2^* > w > w_2^*$. A summary of the conditions for increase of the new gametes for all values of w_1^* and w_2^* is given in Table 5.

There are two cases (IIc2 and IIIc2) in which the new gametes will increase only if the recombination fraction, r, is less than a critical value,

$$\left| \frac{(w-w_2^*)(w_1^*-w)}{(w-w_2^*)vw_{14}-(w_1^*-w)uw_{23}} \right|$$

It is interesting to note, as pointed out by BODMER and PARSONS (1962), that for the symmetrical model considered by LEWONTIN and KOJIMA (1960) in which all homozygotes have the same fitness, $w_1^* = w_2^*$ so that recombination is never critical for the increase of new gametes. Special cases illustrating these conditions will be discussed in a later section. It was also pointed out by BODMER and PARSONS (1962) that, in some cases, when there is a difference between the

TABLE 5

Increase of a gene linked to a polymorphism

$$u = (w_{34}-w_{44})/(2w_{34}-w_{33}-w_{44})$$
$$v = 1-u$$
$$w_1^* = uw_{13} + vw_{14}$$
$$w_2^* = uw_{23} + vw_{24}$$
$$w = uw_{33} + vw_{34} = uw_{34} + vw_{44} = u^2w_{33} + 2uvw_{34} + v^2w_{44}$$
$$r_c = \left| \frac{(w-w_2^*)(w_1^*-w)}{(w-w_2^*)vw_{14} - (w_1^*-w)uw_{23}} \right|$$

		A increases if and only if
I.	$w_1^* = w_2^*$	$w_1^* = w_2^* > w$
II.	$w_1^* > w_2^*$	
a.	$w_1^* > w_2^* > w$	always
b.	$w \geq w_1^* \geq w_2^*$	never
c.	$w_1^* > w \geq w_2^*$	
1.	$(w-w_2^*)vw_{14} \leq (w_1^*-w)uw_{23}$	always
2.	$(w-w_2^*)vw_{14} > (w_1^*-w^*)uw_{23}$	$r < r_c$
III.	$w_2^* > w_1^*$	
a.	$w_2^* > w_1^* > w$	always
b.	$w \geq w_2^* > w_1^*$	never
c.	$w_2^* > w \geq w_1^*$	
1.	$(w-w_1^*)uw_{23} \leq (w_2^*-w)vw_{14}$	always
2.	$(w-w_1^*)uw_{23} > (w_2^*-w)vw_{14}$	$r < r_c$

viability of coupling and repulsion heterozygotes, that is a "position effect," new gametes may only increase in frequency if the recombination fraction between the loci involved is less than some critical value.

The root of equation (33) which has the largest modulus is of some interest since it determines the initial rate of increase of the new gametes, and so, approximately, their ultimate probability of survival when random fluctuations of gene frequencies are taken into account (see BODMER 1960). When $r = 0$, equation (33) has the roots w_1^*/w and w_2^*/w, so that the increase conditions are w_1^*, $w_2^* > w$ corresponding to the conditions for the increase of new alleles at a locus which is already polymorphic for two alleles (see BODMER and PARSONS 1960; HALDANE 1957). When $r/(w_1^* - w_2^*)$ is small and, without loss of generality $w_1^* > w_2^*$, the dominant root of equation (33) takes the approximate form

$$\lambda_1 = (w_1^*/w) - (rvw_{14}/w) + 0(r^2) ,$$

as given by BODMER and PARSONS (1962), and the increase condition is, approximately,

(35) $$r < (w_1^* - w)/vw_{14} .$$

Note that λ_1 increases as the recombination fraction r decreases. When $w_{ij} = 1$ for all i and j, equation (33) has roots $\lambda = 1$ and $1 - r$, as expected from the classical results for two linked loci without selection (JENNING 1917; HALDANE 1926).

8. *Conditions for the Simultaneous Increase of New Alleles at Each of Two Linked Loci*

Consider now a population in which the genotype ab/ab predominates and gametes Ab, aB and AB are all present at low frequencies. We are interested in the conditions under which the increase of the new gametes is dependent on close linkage between the two loci. Let $x_4 = 1 - d$, where $d = x_1 + x_2 + x_3$ and x_1, x_2, x_3, and so d, are all small. BODMER (BODMER and PARSONS 1962), obtained the necessary conditions for this situation by assuming x_1, x_2, x_3, were all of the same order of magnitude and considering the linear equations in these variables, obtained by ignoring quadratic terms, in equations (1). This approach however ignores the fact that, initially, x_1, being the frequency of the gamete AB which contains both new alleles, may be of order $x_2 x_3$, being produced by recombination in the rare double heterozygote Ab/aB. Thus if $x_1 = 0$ initially, $x'_1 = (rw_{23}x_2x_3)/\bar{w}$. Equation (1a) may be written in the form

(45) $$\bar{w}(x'_1 - x_1) = x_1[w_1 - \bar{w} - rw_{14}x_4] + rw_{23}x_2x_3$$

from which it follows that a sufficient condition for $x'_1 > x_1$ for all values of x_1, x_2 and x_3 is $r < (w_1 - \bar{w})/(w_{14}x_4)$ or, since

$$(w_1 - \bar{w})/w_{14} < (w_1 - \bar{w})/w_{14}x_4 \qquad (x_4 \leq 1)$$

(46) $$r < (w_1 - \bar{w})/w_{14} .$$

Now when x_1, x_2 and x_3 are small,

$$w_i = w_{i4} + 0(x_i)$$

so that $\bar{w} = w_{44} + 0(x_i)$ and, to this order of approximation, condition (46) takes the form

(47) $r < (w_{14} - w_{44})/w_{14}$.

Thus, so long as all the x_i are small compared with $(w_{14} - w_{44})/w_{14}$, x_1 will always increase provided condition (47) is satisfied. If we assume $x_1 = 0(x_2 x_3)$ and ignore quadratic terms in x_2 and x_3, and hence in d, equations (1b) and (1c) take the form

(48a) $x_2' = (w_{24}/w_{44})\, x_2$

(48b) $x_3' = (w_{34}/w_{44})\, x_3$.

Thus x_2 and x_3 increase, separately, if and only if $w_{24} > w_{44}$ and $w_{34} > w_{44}$. This follows intuitively from the fact that when gametes Ab and aB are rare, recombinant products from the double heterozygote Ab/aB can be ignored and the gametes can be treated as if they were simply new alleles at the two respective loci. So long as x_2 and x_3 are small, and $x_1 = 0$ $(x_2\, x_3)$, when $w_{24} < w_{44}$, $w_{34} < w_{44}$ and $r < (w_{14} - w_{44})/w_{14}$ it follows from equations (47) and (48) that x_2 and x_3 decrease geometrically while x_1 increases. A state must therefore be reached at which x_1, x_2 and x_3 are all of the same order of magnitude. At this point, the linear approximation considered by BODMER and PARSONS (1962) becomes valid and takes the form

(49a) $w_{44}x_1' = w_{14}(1-r)x_1$

(49b) $w_{44}x_2' = rw_{14}x_1 + w_{24}x_2$

(49c) $w_{44}x_3' = rw_{14}x_1 + w_{34}x_3$

From this it follows that the further increase in x_1 and the subsequent increase in x_2 and x_3 still depend on condition (47). Thus, as indicated by BODMER and PARSONS (1962), when the single heterozygotes (Ab/ab and aB/ab) are less fit than the prevailing homozygote (ab/ab), but the coupling double heterozygote (AB/ab) is fitter than this homozygote, the genes A and B will ultimately increase in frequency if and only if the recombination fraction satisfies condition (47).

TABLE 6

Increase of two new linked genes, A *and* B

	A increases if	B increases if
$w_{14}, w_{24}, w_{34} > w_{44}$	always	always
$w_{14}, w_{24}, > w_{44} \geq w_{34}$	always	$r < r_c^*$
$w_{14}, w_{34} > w_{44} \geq w_{24}$	$r < r_c^*$	always
$w_{14} > w_{44} \geq w_{24}, w_{34}$	$r < r_c$	$r < r_c$
$w_{24}, w_{34} > w_{44} \geq w_{14}$	always	always
$w_{24} > w_{44} \geq w_{14}, w_{34}$	always	never
$w_{34} > w_{44} \geq w_{14}, w_{24}$	never	always
$w_{44} \geq w_{14}, w_{24}, w_{34}$	never	never
	$r_c = (w_{14} - w_{44})/w_{14}$	

* These conditions only apply if the frequency of gamete $AB(x_1)$ is of the same order of magnitude as those of Ab and aB (x_2 and x_3).

When either $w_{24} > w_{44}$ or $w_{34} > w_{44}$ and we start with $x_1 = 0(x_2 x_3)$, it is no longer clear under what circumstances linkage may affect the increase of the new gametes. This problem was discussed by BODMER and PARSONS (1962) using a numerical example. It is clear that whenever x_1 is large enough to be of the same order of magnitude as x_2 and x_3, the conditions derived from equations (49) apply. Then if, for example, $w_{24} > w_{44}$ but $w_{34} < w_{44}$, B will only increase in frequency if condition (47) is satisfied. When, however, x_1 is too small initially, locus A may become polymorphic without locus B. The problem then reduces to that of the conditions for increase of a new gene linked to a polymorphic locus, which has been discussed above. These various conditions for increase are summarized in Table 6.

9. A General Condition for Stable Linkage Disequilibrium When r is Sufficiently Small

When $r = 0$ and $x_2 = x_3 = 0$, gametes AB and ab can be maintained in the balanced polymorphism given by

$$x_1 = u = 1-v = 1-x_4 = (w_{14}-w_{44})/2w_{14}-w_{11}-w_{44})$$

provided $w_{14} > w_{44}$ and w_{11}. Now assume that r is small so that gametes Ab and aB occur with frequencies x_2 and x_3 which are of the same order of magnitude as r. Writing $x_1 = u-d_1$ and $x_4 = v-d_4$, where d_1 and d_4 are also $0(r)$, we have $w_i = uw_{1i} + vw_{4i} + 0(r)$ and $\bar{w} = w + 0(r)$ where $w = u^2 w_{11} + 2uv w_{14} + v^2 w_{44}$ $= uw_{11} + vw_{14} = uw_{14} + vw_{44}$ is the mean fitness of the population when $r = x_2 = x_3 = 0$. Using these approximations, equations (1b) and (1c) take the form

(50a) $$wx_2' = x_2 w_2^* + rw_{14}uv + 0(r^2)$$

(50b) $$wx_3' = x_3 w_3^* + rw_{14}uv + 0(r^2)$$

where $w_2^* = uw_{12} + vw_{24}$ and $w_3^* = uw_{13} + vw_{34}$ are the marginal fitnesses of gametes Ab and aB when first introduced into the population. At equilibrium when $x_2' = x_2$, $x_3' = x_3$ we have

(51) $$x_2 = \frac{rw_{14}uv}{w-w_2^*} + 0(r^2) \qquad \text{and} \qquad x_3 = \frac{rw_{14}uv}{w-w_3^*} + 0(r^2)$$

which are valid equilibria provided $w > w_2^*$ and w_3^* respectively. This corresponds to the equilibrium (22a) given by the symmetrical viability model when r is small. Writing equation (50a) in the form

$$w(x_2'-x_2) = x_2(w_2^*-w) + rw_{14}uv + 0(r^2)$$

it is clear that when $w_2^* > w$, x_2 always increases for sufficiently small r and similarly for x_3 when $w_3^* > w$. Thus when w_2^*, $w_3^* > w$ no stable equilibrium can exist with both x_2 and x_3 of $0(r)$, so that $w > w_2^*$, w_3^* is a necessary condition for the stability as well as the existence of the equilibrium given by equation (51). An exactly analogous equilibrium exists in the neighborhood of $x_1 = x_4 = 0$ and

$$x_2 = 1 - x_3 = \frac{w_{23} - w_{33}}{2w_{23} - w_{22} - w_{33}}$$
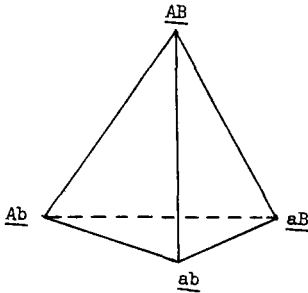
when r is sufficiently small.

FIGURE 1.—Tetrahedral representation of gamete frequencies. A point inside the tetrahedron represents a population whose respective gamete frequencies are the perpendicular distances from the four faces of the tetrahedron.

10. *Sufficient Conditions for the Existence of a Two-Locus Polymorphism*

The gametic frequencies of a population in which two loci are each segregating for two alleles can be represented as a point in an equilateral tetrahedron such that the distances from the four faces of the tetrahedron are proportional to the frequencies of the four gamete types AB, Ab,, aB, and ab. The sum of the four distances will equal the altitude of the tetrahedron, no matter where the point is placed (Figure 1).

A population fixed for one of the four gamete types will lie at one of the four vertices, while a population which has lost one allele at one of the loci will lie on one of the four edges AB-aB, AB-Ab, Ab-ab or ab-aB. We wish to find conditions such that the population can ever remain at a vertex, an edge, or a face, so that it must ultimately come to rest in the interior of the tetrahedron, where both loci are segregating. If we assume $r > 0$, then the system cannot have an equilibrium point on either of the edges AB-ab or aB-Ab, for recombination will continually produce the missing gamete types. Likewise there cannot be an equilibrium on any of the faces of the tetrahedron, since a population located on a face lacks only one gamete type. Recombination will produce this missing gamete, forcing the population into the interior of the tetrahedron. This leaves four edges and the four vertices. Each of the vertices is automatically an equilibrium point if there is no mutation, since at each of these points only one gamete type exists and this situation admits of no change by selection or genetic drift. We can determine whether the equilibria at the four vertices are stable by using the conditions derived above (Table 6) for the simultaneous increase of two new alleles at each of two loci. If the missing alleles can increase in frequency when introduced at low frequency at either or both loci, the equilibrium cannot be stable. If introduced alleles always disappear (in an infinite population) then the equilibrium is stable.

If an equilibrium point exists on one of the four edges (AB-ab, aB-ab, AB-aB, and Ab-ab), it will be a selectively balanced polymorphism at one of the two loci maintained by overdominance. For such an equilibrium we can use the conditions given in Table 5 to determine whether the missing allele at the fixed locus will increase if introduced at a low frequency, and hence whether the equilibrium is stable. The stability of the equilibria near the remaining two edges AB-ab and Ab-aB can also be investigated using the conditions derived in the previous section for sufficiently small values of r.

Thus, given any set of fitnesses and the value of r ($\neq 0$), we can determine whether any equilibrium point located on an edge or at a vertex is stable. If all such points are unstable, then if there is a stable equilibrium point at which all four gamete types exist and hence both loci are segregating, it must be somewhere in the interior of the tetrahedron.

We shall first apply this type of analysis in turn to the additive, multiplicative and general symmetric viability models given in Tables 2, 3 and 4 and then we shall discuss in detail an illustrative numerical example.

Consider first the additive model given in Table 2. It was shown above that a unique internal solution with $D = 0$ exists provided the pairs of differences $a_2 - a_1$, $a_2 - a_3$ and $b_2 - b_1$, $b_2 - b_3$ have the same sign. If these differences all have positive signs, then all four vertices are unstable equilibrium points, and vice versa if all are negative. All four vertices are also unstable if only two of the differences are negative. Now consider the application of the conditions given in Table 5 to this model. We have

$$w_1^* = a_2 + ub_2 + vb_3, \qquad w_2^* = a_2 + ub_1 + vb_2$$
$$w = a_3 + u(ub_1 + vb_2) + v(ub_2 + vb_3)$$

where

$$u = 1 - v = (b_2 - b_3)/(2b_2 - b_1 - b_3)$$

Therefore $w_1^* - w = w_2^* - w = a_2 - a_3$ and so the equilibrium on the edge aB-ab is stable if $a_2 < a_3$ and unstable if $a_2 > a_3$. Similarly the equilibria on the edges AB-aB, Ab-Ab, and Ab-ab are unstable if $b_2 > b_1$, $a_2 > a_1$ and $b_2 > b_3$, respectively. Thus, in summary

i. If $a_2 - a_1$, $a_2 - a_3$, $b_2 - b_1$, $b_2 - b_3$ are all positive, the only stable equilibrium is the internal point at which $D = 0$.

ii. If, for example, $a_2 - a_1$, $a_2 - a_3$, $b_2 - b_1$ are all positive but $b_2 - b_3$ is negative then the only stable equilibria are the vertices aB and ab.

iii. If $a_2 - a_1$, $a_2 - a_3$ are positive and $b_2 - b_1$, $b_2 - b_3$ are negative, then both edges Ab-aB and Ab-ab have stable equilibrium but no vertices or internal points are stable.

iv. If, for example, $a_2 - a_1$ and $b_2 - b_1$ are positive, while $a_2 - a_3$ and $b_2 - b_3$ are negative, then only the vertex ab is stable.

v. If, for example, $a_2 - a_1$ is positive and $a_2 - a_3$, $b_2 - b_1$, $b_2 - b_3$ are all negative then the only stable equilibrium are the vertices aB and ab.

Lastly

vi. If all of $a_2 - a_1$, etc. are negative all four vertices are stable and no other stable equilibrium points exist.

Clearly, where there is more than one stable equilibrium, which one is reached will depend on the initial configuration of the population. For the additive model, which is undoubtedly the simplest to analyze, linkage has no effect on the equilibrium configurations though it may affect the rate of approach to the various equilibria.

In the multiplicative model shown in Table 3, we can without loss of generality write $\alpha_1 = 1 - s_1$, $\alpha_2 = 1$, $\alpha_3 = 1 - t_1$; $\beta_1 = 1 - s_2$, $\beta_2 = 1$, $\beta_3 = 1 - t_2$, to give the parametrization shown in Table 7. The internal equilibrium point is then given by

TABLE 7

An alternative fitness scheme for the multiplicative model

	BB	Bb	bb
AA	$(1-s_1)(1-s_2)$	$(1-s_1)$	$(1-s_1)(1-t_2)$
Aa	$(1-s_2)$	1	$(1-t_2)$
aa	$(1-t_1)(1-s_2)$	$(1-t_1)$	$(1-t_1)(1-t_2)$

$$p_1 = t_1/(s_1+t_1), \qquad p_2 = t_2/(s_2+t_2)$$

and $D = 0$, and is valid only if s_1 and t_1, and s_2 and t_2 have the same signs. From the analysis of the symmetric model and from the analysis of equilibria near the edges *AB-ab* and *Ab-aB* (equations (51)), it follows that, in general, there exist two further internal equilibria for which $D \neq 0$. Equation (14) gives a necessary condition for an equilibrium with $D \neq 0$ to be stable. Consider now the conditions for the stability of the equilibrium on the edge *aB-ab*, which only exists if s_2, $t_2 > 0$. We have

$$u = 1 - v = t_2/(s_2+t_2)$$
$$w_1{}^* = w_2{}^* = 1 - s_2t_2/(s_2+t_2) \qquad \text{and} \qquad w = (1-t_1)(1 - s_2t_2/(s_2+t_2)).$$

Thus, from Table 5, the equilibrium is stable if $t_1 < 0$ so that $w_1{}^* = w_2{}^* < w$, and is unstable if $t_1 > 0$ so that $w_1{}^* = w_2{}^* > w$, the stability being independent of the value of r. The conditions for the stability of the edges *AB-Ab*, *Ab-ab*, *AB-aB* are, therefore, $s_1 < 0$, $t_2 < 0$ and $s_2 < 0$, respectively. If at least three of s_1, t_1, s_2, t_2 are positive then all vertices are unstable. If two of these quantities are positive, then either three or four vertices are unstable depending on which two are positive, while if only one is positive, just two vertices are unstable. All vertices are stable only when all of s_1, t_1, s_2 and t_2 are negative. In summary

i. If all of s_1, t_1, s_2, t_2 are positive, the equilibrium with $D = 0$ is the only stable equilibrium if

$$r > \left(\frac{s_1t_1}{s_1 + t_1} \right) \left(\frac{s_2t_2}{s_2 + t_2} \right)$$

but if r is less than this critical value one or more internal points with $D \neq 0$ will be stable.

ii. If, for example, only $t_2 < 0$ then the only marginal equilibrium which is stable is that on the edge *Ab-ab*. The equilibrium with $D = 0$ does not exist, but one or two internal equilibria may exist which are stable for sufficiently small r.

iii. If, for example, s_1 and t_1 are positive but s_2 and t_2 are negative, then the equilibria on the edges *AB-aB* and *Ab-ab* are stable but none of the vertices are stable. The equilibrium with $D = 0$ is presumably unstable.

iv. If, for example, s_1 and t_2 are positive but s_2 and t_1 are negative, then only the vertex *aB* is stable on the margin, and the equilibrium $D = 0$ does not exist.

v. If only one of s_1, etc is positive, say s_1, and the others are negative then two of the vertices (*AB* and *Ab*) are stable, but no other marginals points are stable.

Lastly

vi. If all of s_1, etc. are negative, then, presumably, the only stable points are the four vertices.

It is not clear without further analysis whether or not in cases (ii), (iii), (iv) and (v) stable internal equilibrium points with $D \neq 0$ may exist for sufficiently small r.

We next consider the application of the conditions for stability of the marginal equilibria to the general symmetrical model given in Table 4. The vertices AB and ab will be unstable if either γ or $\beta < \delta$ or if when $\beta, \gamma > \delta$ the recombination fraction $r < \delta$. Similar stability conditions apply for the vertices Ab and ab but with α replacing δ. The equilibrium on the edge $aB\text{-}ab$ occurs at $u = 1 - v = (\delta - \beta)/(\alpha + \delta - 2\beta)$ and exists and is stable, in the absence of gametes AB, Ab if and only if α and $\delta > \beta$. When this equilibrium exists we have, following

$$w_1{}^* = 1 - \frac{\gamma(\delta - \beta)}{\alpha + \delta - 2\beta}, \qquad w_2{}^* = 1 - \frac{\gamma(\alpha - \beta)}{\alpha + \delta - 2\beta}$$

$$\text{and} \qquad w = 1 - \frac{(\alpha - \beta)\,(\delta - \beta)}{\alpha + \delta - 2\beta}$$

Thus $w_1{}^* - w$ and $w_2{}^* - w$ are both positive if $\alpha > \beta + \gamma$ and $\delta > \beta + \gamma$ respectively, in which case the equilibrium is unstable for all r, while if $\alpha, \delta < \beta + \gamma$ it is stable for all r. When, however, say, $\alpha > \beta + \gamma > \delta$, the value of r may determine the stability of the equilibrium following the conditions (2c) and (3c) given in Table 5. By symmetry, these same conditions apply to the edges $AB\text{-}Ab$, $Ab\text{-}ab$ and $AB\text{-}aB$. The equilibria near the edges $AB\text{-}aB$ and $Ab\text{-}aB$ correspond, as indicated above, to the equilibria 22a and 22b for small r. These only exist if $\beta + \gamma > \delta$ and $\beta + \gamma > \alpha$ respectively. It is clearly not feasible to enumerate all the possible equilibrium configurations. One or two examples are, however, worth discussing. Thus suppose, for example, that $\beta, \gamma > \delta > 0$ while $\alpha > \beta + \gamma$ so that if $r < \delta$, all the vertices are unstable. Then, provided r is sufficiently small, a stable internal equilibrium with $D \neq 0$ will exist in the neighborhood of the edge $AB\text{-}ab$. In this case the value of r is critical for the determination of the overall equilibrium configuration. If $\beta + \gamma < \delta$ and α, and $\beta, \gamma > 0$, no vertices or edges are stable and the equilibrium near $D = 0$ is the only one which is stable.

We shall now use the set of fitnesses in Table 8 as the basis for a more complete numerical illustration of the application of these conditions for the stability of the marginal equilibria. We first examine the stability of the four vertices of the tetrahedron. For AB, the set of fitnesses to be tested is $w_{44} = .70$, $w_{24} = w_{34} = .9$ and $w_{14} = 1.0$ corresponding to the upper left hand corner of Table 8. Note that, for comparison with the formulation of the section on the simultaneous increase

TABLE 8

Fitness scheme for illustrative numerical example

	BB	Bb	bb
AA	.70	.90	.80
Aa	.90	1.00	.40
aa	.80	.40	.90

of two new alleles at two linked loci, it is necessary to renumber the gametes so that the double homozygote fitness is w_{44}, and the single homozygote fitnesses are w_{24} and w_{34}, while the double heterozygote fitness is w_{14}. Since w_{14}, w_{24} and $w_{34} > w_{44}$, a and b always increase so that the equilibrium at the vertex AB is unstable for all values of r.

For the Ab equilibrium, $w_{14} = 1.0$, $w_{24} = .4$, $w_{34} = .9$, and $w_{44} = .8$. So w_{14}, $w_{34} > w_{44} > w_{24}$ and B will always increase. Therefore the equilibrium is unstable. Since we seek only conditions under which at least one of the fixed alleles increases when both are introduced at low frequency, the condition on r is irrelevant to our argument. The aB equilibrium also is always unstable, since the fitnesses are symmetrical when A is exchanged for a and simultaneously B is exchanged for b.

The ab equilibrium has $w_{14} = 1.0$, $w_{24} = w_{34} = .4$, and $w_{44} = .9$. Then $w_{14} > w_{44} \geq w_{24}$, w_{34} and so from condition (47), alleles A and B increase only when $r < (1.00 - .90)/1.00 = .10$. If $r = 0.1$, the equilibrium is neutral, and if $r > 0.1$, it is stable. The foregoing arguments tell us that we are so far assured of an interior stable equilibrium only when $r < 0.1$. Examining the possibilities of a stable equilibrium with one locus segregating, we see that if a is fixed and locus B segregates, the relevant fitnesses are $w_{33} = .8$, $w_{34} = .4$, and $w_{44} = .9$. This set of fitnesses shows underdominance so that a stable equilibrium could not exist on the edge aB-ab. Likewise, by symmetry, no stable equilibrium is possible with b fixed and locus A segregating. If A is fixed and locus B segregates we have, renumbering gamete types,

$$w_{33} = .7, \qquad w_{34} = .9, \qquad w_{14} = .8,$$
$$w_{13} = .9, \qquad w_{14} = w_{23} = 1.00, \qquad w_{24} = .8.$$

The equilibrium frequency at locus B when A is fixed is

$$u = (w_{34} - w_{44})/(2w_{34} - w_{33} - w_{44}) = .333$$
$$v = 1 - u = .666$$

At this point

$$w_1^* = uw_{13} + vw_{14} = .966$$
$$w_2^* = uw_{23} + vw_{24} = .600$$
$$w^* = uw_{33} + vw_{34} = .833$$

Referring to Table 5, we find that $w_1^* > w_2^*$ and $w_1^* > w \geq w_2^*$ so that we must calculate $(w-w_2^*)vw_{14}$ and $(w_1^*-w)uw_{14}$. These are

$$(w-w_2^*)vw_{14} = .155 \text{ and } (w_1^*-w)uw_{23} = .044$$

since $(w-w_2^*)vw_{14} > (w_1^*-w)uw_{23}$ allele a will increase whenever

$$r < \frac{(.233)(.133)}{(.155)-(0.44)} = .280.$$

Since the equilibrium in which B is fixed and locus A segregates is symmetrical with this equilibrium, the same condition holds in that case. We have already seen that when, the population is fixed for genotype $aabb$, that equilibrium is unstable only when $r < .1$. So, when $r < .1$, all equilibria at which one locus is fixed or both loci are fixed are unstable. Any population started close to the

boundary of the tetrahedron will then be pushed by selection into the interior, where there must be a stable equilibrium with both loci segregating.

A computer program has been written to check the validity of the above calculations. Using equations (1), the program produces successive generations in a hypothetical infinite population. The fitnesses of Table 8 were used, and r varied from .05 to .5 in steps of .05. In one set of runs allele B was started at a frequency of .33 and allele a at a frequency of .01, all of the a alleles being in ab gametes in one set of runs and in aB gametes in another. This corresponds fairly closely to a situation in which locus B would be segregating at frequencies maintained by selection, with a small proportion of ab or aB gametes being introduced into the population. The results confirmed the above calculations. When r was .05, .1, .15, .2 or .25, allele a ultimately increased in the population, although if it was introduced as ab gametes it would temporarily decrease to start with. When r was .3, .35, .4, .45, or .5, allele a ultimately decreased, although if it was introduced by aB gametes it would temporarily increase to start with. This demonstrates that the above approximations can hold if the initial frequency of the rare allele is as high as .01, the approximations thus being reasonably robust.

The other set of runs started with alleles A and B at initial frequencies of .01. In half of the runs all of the A and B alleles were initially in AB gametes while in the other half they were initially in Ab and aB gametes. Again the runs showed that the above calculations are valid. When r was .05, both alleles increased. When r was .15 or greater, both decreased. When r was .1, both alleles would have been expected to remain at their initial frequencies or decrease, since the theoretical condition for increase is $r < .1$. In the computer runs, A and B remained almost unchanged after some initial change. In one case (A and B initially in coupling) there was a slow increase of A and B. This is interpreted as due to second-order effects, since the frequency of A and B was not infinitely small but was .01, so that some genotypes ignored in the above calculations, such as AB/AB, actually existed in the population at low frequencies and affected the rate of change of A and B. The computer was also used to solve for stable equilibria of the gamete frequencies. Stable equilibria were found at all values of r. This raises the point that the conditions given above for the existence of stable equilibria with both loci segregating are *sufficient* conditions, but not *necessary* conditions. For example, when $r = .5$, there was a stable equilibrium at $x_1 = .4762$, $x_2 = x_3 = .2251$, and $x_4 = .0736$.

It is interesting to note that any set of fitnesses in which the double heterozygotes are more fit than the corresponding single heterozygotes, and the single heterozygotes are more fit than the corresponding double homozygotes will always have a stable interior equilibrium. Reference to Table 5 will show that we always have w_1^*, $w_2^* > w$ and also w_{14}, w_{24}, $w_{34} > w_{44}$ so that introduced alleles always increase if $r > 0$.

<div align="center">DISCUSSION</div>

The essence of our analysis of the two locus model with linkage and selection is to identify conditions under which linkage causes a significant departure from

the gametic changes and equilibria which are expected in the absence of selection, when $D = 0$. There are two major aspects of this analysis. On the one hand we have attempted to derive conditions for the existence of stable equilibria for which D is appreciably different from 0 and on the other hand we have identified conditions under which close linkage plays a critical role in the increase of new gametic combinations. The validity of the results of the analysis of any model is, of course, bounded by the assumptions built into the model. While more complex two locus models may certainly give results differing quantitatively from those presented here, they will not alter the qualitative conclusion that plausible situations do exist for which linkage can profoundly affect the selection and ultimate equilibrium of new gametic combinations.

In all our discussion of stability we have assumed that an equilibrium is effectively unstable provided at least one of the eigenvalues for linear perturbation about the equilibrium has modulus greater than unity. Mathematically this is not strictly correct, since, in general, this only ensures movement away from the equilibrium in certain directions, depending on the eigenvector corresponding to the maximum eigenvalue. Instability *in all directions* is only assured if *all* eigenvalues have modulus greater than unity. In reality, however, random fluctuation in gametic frequencies always occur, and are bound, sooner or later, to bring the gamete frequencies into a region of instability, even when there is only one eigenvalue with modulus greater than unity. When this happens, the gamete frequencies will nearly always move away from the equilibrium by a definite positive amount and so the equilibrium is, for practical purposes, unstable (for further discussion see BODMER, FELDMAN and KARLIN (in preparation).

The analyses presented here have been deterministic, having ignored in particular random fluctuations due to very low gene frequencies and due to finite population size. As already pointed out, the initial rate of selection of the new gamete provides, following FISHER (1930), at least an approximate measure of the ultimate chance of survival of a single new occurrence. A general stochastic treatment of the two locus model, even in its simplest form, is clearly very difficult, though some useful results have been obtained by KOJIMA and SCHAFER (1964), LATTER (1966), KARLIN, McGREGOR and BODMER (1967), and HILL and ROBERTSON (1966).

It is perhaps of some interest to conjecture qualitatively, the likely effect of drift due to small finite population sizes. For each stable equilibrium point there will be a region of the tetrahedron within which selection will tend to move a population toward the equilibrium. Consider a case in which there is a stable equilibrium in the interior of the tetrahedron and also a stable equilibrium at an edge or vertex. Normally, random genetic drift will result in the population wandering about in the region of the interior equilibrium. Occasionally, however, the population will be carried into the region corresponding to the equilibrium on the boundary of the tetrahedron. Selection will then tend to pull the population to the edge or vertex, resulting in fixation at one or both loci. If the missing allele or alleles are subsequently produced by mutation or introduced by immigration, this will result in displacement of the population from the edge or vertex by a

small amount. The population will still be within the "region of attraction" of the boundary equilibrium, so that if it moves to the interior equilibrium, it will be as a result of random genetic drift opposing the selection pressure.

On the other hand, if all of the boundary equilibria are unstable (so that there must exist a stable interior equilibrium), the population will become fixed at one or both loci less readily, since selection pressure will tend to return the population to the interior equilibrium. When a fixed population becomes unfixed as a result of mutation or immigration, the population will have been moved off the boundary into the region in which selection tends to pull the population toward the interior equilibrium. If the population then becomes fixed again, it will be in spite of the selection, whereas in the previous case selection promoted fixation. Thus if the boundary equilibria are unstable the population will spend more of its time in the interior of the tetrahedron than it otherwise would. The condition that no marginal point be a stable equilibrium is therefore, in a sense *necessary for the long-term maintenance* of a two-locus polymorphism in a finite population.

The extension of the two locus analysis to more loci is especially important. LEWONTIN (1964b) has obtained very interesting results by computer simulation, showing how the interaction of linkage and selection can maintain relatively large blocks of genes together on a chromosome, provided each pair of adjacent genes satisfies the conditions for linkage disequilibrium predicted by the two locus models.

The evidence for evolutionary molding of the organization of genes on chromosomes was extensively reviewed by BODMER and PARSONS (1962) and will not be further discussed here. It is of course our hope, that the results presented here provide some indication of the type of selection needed to effect evolutionary modification of the positioning of genes on the chromosomes. It seems worth emphasizing that, at least in higher organisms, an interval on the chromosome corresponding to a recombination fraction of one or a few percent may contain many, perhaps even hundreds, of cistrons. Our results and those of others, suggest that recombination fractions of this order of magnitude may often cause very significant departures from random association of genes on the chromosomes, though larger recombination fractions are less likely to have appreciable qualitative effects. Thus, clusters of genes which are relatively close together may often be maintained in their relative positions by the interaction between selection and linkage. Genes which are genetically far apart, are, however, much less likely to be maintained in their given relative positions by such selective interactions. In other words, the effect of the interaction between linkage and selection is one that is expected to be concentrated in short regions of the genome and does not necessarily extend over large portions of any particular chromosome.

SUMMARY

The mathematical analysis of the deterministic two locus random mating model with linkage and selection is reviewed and extended following mainly the

previous analyses by WRIGHT (1952), KIMURA (1956), LEWONTIN and KOJIMA (1960) and BODMER (BODMER and PARSONS 1962)— —1. The general equations (1) for the model are first derived.— —2. A condition on the fitnesses for the existence of a non-trivial equilibrium with $D = 0$ is next derived (equation (6)).— — 3. An "additive" fitness model leads to a unique internal equilibrium with $D = 0$, which is the unique stable equilibrium if both loci are heterotic.— —4. A "multiplicative" model also gives an equilibrium with $D = 0$, which, however, in general is stable only if r exceeds a simple function of the selection coefficients. When r is less than its critical value, stable equilibria with $D \neq 0$ may exist.— —5. A general necessary condition (13) for the stability of a nontrivial equilibrium with $D = 0$ is derived. The application of the condition to the additive and multiplicative (14) models is discussed.— —6. A general symmetrical viability model, which includes essentially all those previously considered, is analyzed in detail for equilibria and their stability. As for the multiplicative model, equilibria with D appreciably different from zero may exist provided r is less than a certain critical value.— —7, 8. Conditions are derived for the increase of a gene linked to a stable polymorphism (Table 5) and for the simultaneous increase of two new alleles at each of two linked loci (Table 6).— —9. A general condition is derived for the existence of a stable linkage disequilibrium whenever r is sufficiently small.— —10. The overall pattern of equilibria given by these models as a function of fitnesses and the recombination fraction is reviewed with reference to the representation of the gamete frequencies of a population as a point in a regular tetrahedron. A sufficient condition for the existence of a stable internal equilibrium point is that no marginal point (either on an edge or a vertex) is a stable equilibrium point. The conditions for the increase of new gametic combinations provide the criteria for the stability of the marginal points of the tetrahedron. The additive, multiplicative and symmetric viability models as well as a special numerical example are used to illustrate these general principals.— —It is emphasized in the discussion that the reason for an interest in the detailed analysis of the interaction between linkage and selection, is in order to obtain some indication of the type of selection which is needed to effect the evolutionary molding of the organization of genes on the chromosomes.

LITERATURE CITED

BENNETT, J. H., 1965 Estimation of the frequencies of linked gene pairs in random mating populations. Am. J. Hum. Genet. **17**: 51–53.

BODMER, W. F., 1960 Discrete stochastic processes in population genetics. J. Roy. Stat. Soc. B **22**: 218–236. —— 1965 Differential fertility in population genetics models. Genetics **51**: 411–424.

BODMER, W. F., and P. A. PARSONS, 1960 The initial progress of new genes with various genetic systems. Heredity **15**: 283–299. —— 1962 Linkage and recombination in evolution. Advan. Genet. **11**: 1–100.

BODMER, W. F., and R. PAYNE, 1965 Theoretical consideration of leukocyte grouping using multispecific sera. pp. 141–149. *Histocompatibility Testing 1965*. Series Hoemat., 11 Munksgaard, Copenhagen.

BODMER, W. F., J. BODMER, S. ADLER, R. PAYNE, and J. BIALEK, 1966   Genetics of "4" and "LA" human leukocyte groups. Ann. N.Y. Acad. Sci. **129**: 473–489.

CROW, J. F., and M. KIMURA, 1965   Evolution in sexual and asexual populations. Am. Naturalist **99**: 439–450.

EWENS, W. J., 1966   Linkage and the evolution of dominance. Heredity **00**: 363–370. ——— 1967   The probability of fixation of a mutant; the two locus case. (in press).

FELSENSTEIN, J., 1965   The effect of linkage on directional selection. Genetics **52**: 349–363.

FISHER, R. A., 1930   *The Genetical Theory of Natural Selection.* Oxford University Press. (2nd edition 1958, Dover).

GIBSON, J. B., and J. M. THODAY, 1959   Recombinational lethals in a polymorphic population. Nature **184**: 1593–1594.

HALDANE, J. B. S., 1926   Mathematical theory of artificial and natural selection. III. Proc. Camb. Phil. Soc. **23**: 363–372. ——— 1957   The conditions for coadaptation in polymorphism for inversions. J. Genet. **55**: 218–225. ——— 1963   The selection of double heterozygotes. J. Genet. **58**: 125–128.

HALDANE, J. B. S., and S. D. JAYAKAR, 1963   The elimination of double dominants in large random mating populations. J. Genet. **58**: 243–251.

HILL, W. G., and A. ROBERTSON, 1966   The effect of linkage on limits to artificial selection. Genet. Res. **8**: 269–294.

JAIN, S. K., and R. W. ALLARD, 1965   The nature and stability of equilibria under optimizing selection. Proc. Natl. Acad. Sci. U.S. **54**: 1436–1443. ——— 1966   The effects of linkage, epistasis and inbreeding on population changes under selection. Genetics **53**: 633–659.

JENNINGS, H. S., 1917   The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. Genetics **2**: 97–154.

KARLIN, S., J. McGREGOR, and W. F. BODMER, 1967   The rate of production of recombinants between linked genes in finite populations. Proc. 5th Berkeley Symp. Prob. Stat. (in press).

KIMURA, M., 1956   A model of a genetic system which leads to closer linkage by natural selection. Evolution **10**: 278–287. ——— 1965   Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. Genetics **52**: 875–890.

KOJIMA, K., 1965   The evolutionary dynamics of two-gene systems. Chapter 9, pp. 197–220. *Computers in Biomedical Research*, Vol. 1.

KOJIMA, K., and T. M. KELLEHER, 1961   Changes of mean fitness in random mating population when epistasis and linkage are present. Genetics **46**: 527–540.

KOJIMA, K., and H. E. SCHAFER, 1964   Accumulation of epistatic gene complexes. Evolution **18**: 127–129.

LATTER, B. D. H., 1966   The interaction between effective population size and linkage intensity under artificial selection. Genet. Res. **7**: 313–323.

LEWONTIN, R. C., 1964a   The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49**: 49–67. ——— 1964b   The interaction of selection and linkage. II. Optimum models. Genetics **50**: 757–782. ——— 1964c   The role of linkage in natural selection. Proc. 11th Intern. Congr. Genet. **3**: 517–525.

LEWONTIN, R. C., and K. KOJIMA, 1960   The evolutionary dynamics of complex polymorphisms. Evolution **14**: 458–472.

MATHER, K., 1941   Variation and selection of polygenic characters. J. Genet. **41**: 159–193.

MORAN, P. A. P., 1964   On the nonexistence of adaptive topographies. Ann. Hum. Genet. **27**: 383–393. ——— 1967   Unsolved problems in Evolutionary theory. Proc. 5th Berkeley Symp. Prob. Stat. (in press).

MORTON, N. F., and M. P. MI, 1966   Blood factor association. Vox Sanguinis **11**: 434–449.

MORTON, N. F., H. KRIEGER, and M. P. MI, 1966   **Natural selection on polymorphisms in** Northeastern Brazil, Am. J. Hum. Genet. **18**: 153–171.

NEI, M., 1964a   Effects of linkage and epistasis on the equilibrium frequencies of lethal genes. I. Linkage equilibrium. Japan J. Genet. **39**: 1–6. —— 1964b   Effects of linkage and epistasis on the equilibrium frequencies of lethal genes. II. Numerical solutions. Japan J. Genet. **39**: 7–25.

PARSONS, P. A., 1963a   Complex polymorphisms where the coupling and repulsion double heterozygote viabilities differ. Heredity **18**: 369–374. —— 1963b   Polymorphisms and the balanced polygenic combination. Evolution **17**: 564–574.

SINGH, M., and LEWONTIN, R. C., 1966   Stable equilibria under optimizing selection. Proc. Natl. Acad. Sci. U.S. **56**: 1345–1348.

STAHL, F. W., and N. MURRAY, 1966   The evolution of gene clusters and genetic circularity in microorganisms. Genetics **53**: 569–576.

WRIGHT, S., 1952   The genetics of quantitative variability. pp. 5–41. *Quantitative Inheritance.* Edited by K. MATHER. London, Her Majesty's Stationary Office. —— 1965   Factor interaction and linkage in evolution. Proc. Roy. Soc. London B **162**: 80–104.