# STUDIES ON ARTIFICIAL SELECTION

by

WILLIAM GEORGE HILL

B.Sc. (London)

M.S. (California)

Thesis presented for the Degree of Doctor of Philosophy

of the University of Edinburgh in the Faculty of Science.


Institute of Animal Genetics,
Edinburgh.

October, 1965.

C O N T E N T S

CONTENTS (continued)



APPENDICES

Papers submitted for publication under joint authorship:

## 1. INTRODUCTION

Although all populations undergoing artificial selection must be of finite size, most of the theory of long term response to artificial selection has been developed for infinitely large populations. Algebraic difficulties have made it necessary for much of the theory, even of infinite populations, to be developed for single genes and only recently has linkage been included (Griffing, 1960). Furthermore, it has generally been assumed that individual genes have effects sufficiently small that changes in genetic parameters, other than the population mean, can be ignored. Using a model of two loci in an infinite population Nei (1963) and Felsenstein (1965) have developed formulae for the effect of directional selection on changes in linkage disequilibrium and selection response. Nei (1963) also derived equations for the expected changes in the components of genetic variance each generation and gave some numerical results for long term predictions.

Unless there is heterozygote superiority all favourable genes will eventually be fixed if the population is infinitely large. However, in small populations, favourable genes may be lost by chance so that predictions are needed not only for the rate of selection advance but also the selection limit. Robertson (1960) introduced a theory of limits to artificial selection in small populations in terms of single genes, which he extended to multiple loci by ignoring linkage and epistatic interactions between loci. Simulation by Monte Carlo methods on a high speed computer has shown that, although populations may initially be in linkage equilibrium, the selection

limit is reduced when genes are tightly linked, even with non-epistatic loci (Martin and Cockerham, 1960; Gill, 1963; Qureshi, 1963). However, these workers all used models in which each gene had the same effect and initial frequency one-half so that generalised conclusions on the role of linked genes were not obtained. Linkage has been shown to have a more marked influence on selection response and limits in small populations that are initially in disequilibrium (Fraser, 1957b; Martin and Cockerham, 1960).

Information on the effects of linkage on artificial selection limits may be of use in designing selection experiments and commercial breeding programmes, so this study was undertaken to extend Robertson's (1960) theory to include some aspects of linkage. Most of the investigation is in terms of only two loci, each with two alternative alleles. Whilst this must greatly oversimplify the situation in nature, such a simple model allows a more thorough analysis of the effects and interactions of the various parameters. Even with two loci each of two alleles sixteen parameters could be considered: three degrees of freedom amongst the frequencies of the four types of gamete, ten genotypes each with a different selective value, the population size and the recombination fraction between the loci. Since this number of variables could not be handled in detail, a simple model was used in which only additive genes and populations in initial linkage equilibrium were included. An explicit general solution even for the additive model could not be found, so that most of the results have been obtained by Monte Carlo simulation.

## 2. RESULTS FROM INFINITE POPULATION

In this section the notation is introduced and some effects of linkage on response to artificial selection in infinitely large random mating populations are discussed. These results will form a basis for the small population study which follows.

Two loci

Let the two loci each have two alternative alleles A, a and B, b where the alleles A and B are taken to have a favourable effect on some trait. Let

$p$ be the frequency of the allele A,

$q$ B,

$f_1$ gamete AB,

$f_2$ Ab,

$f_3$ aB,

$f_4$ ab,

$c$ be the recombination fraction between the loci and be the same for both sexes. Linkage disequilibrium will be measured by $\Delta$, the linkage disequilibrium determinant, where

$$\Delta = f_1 f_4 - f_2 f_3 .$$

Positive values of $\Delta$ imply an excess of coupling heterozygotes (AB, ab) and negative values an excess of repulsion heterozygotes (Ab, aB) beyond their frequency expected from independent association of their constituent genes. It can be shown that

$$\left.\begin{aligned} f_1 &= pq + \Delta , & f_2 &= p(1-q) - \Delta , \\ f_3 &= (1-p)q - \Delta , & f_4 &= (1-p)(1-q) + \Delta . \end{aligned}\right\} \quad (1)$$

The largest positive value which the disequilibrium determinant can take is $p(1-q)$ or $(1-p)q$, whichever is smaller, and the largest negative value it can take is $pq$ or $(1-p)(1-q)$, whichever is smaller (Lewontin, 1964).

If $\Delta_t$ is the determinant at generation t, then in a large random mating population in which there are discrete generations and no selection, linkage disequilibrium is reduced at the rate

$$\Delta_t = (1-c)\,\Delta_{t-1}$$

$$= (1-c)^t \Delta_o .$$

Let the genotypic value of the zygote formed from the gametes with frequencies $f_j$ and $f_k$ be $v_{jk}$ for some trait of interest and let $w_{jk}$ be its selective advantage. $w_{jk}$ is defined as the probability that an individual with genotypic value $v_{jk}$ is selected as a proportion of the probability that an individual taken at random from the population is selected. The latter is, of course, the fraction of the population selected as parents of the next generation. With random mating, the frequency of a genotype is the product of the frequencies of its constituent gametes. Thus, from the definition of selective advantage used, it follows that $\sum_j \sum_k f_j f_k w_{jk} = 1$.

If truncation selection is practised on the individual phenotype the selective advantage of a genotype is given by

$$w_{jk} = 1 + \frac{i}{\sigma}(v_{jk} - m) \qquad (2)$$

where m is the population mean and i the selection differential in standard deviations (Kimura, 1958; Griffing, 1960). For the derivation of equation (2) it is assumed that gene effects are small

relative to the phenotypic standard deviation ($\sigma$), the selection intensity is low and phenotypes are normally distributed. Latter (1965) has investigated the consequences of departures from these assumptions. He found that the relative probability is poorly estimated by (2) when less than about 40% of the population is selected and gene effects are such that $|(v_{jk} - m)/\sigma| \geq 0.5$. However changes in gene and gametic frequency are less seriously affected, for with intense selection (2) underestimates the relative probability of selection for both positive and negative values of $(v_{jk} - m)/\sigma$.

The change in gametic frequency, obtained by modifying a formula of Lewontin and Kojima (1960) to the notation used here, is for one generation of selection

$$df_j = f_j \left(\sum_k f_k w_{jk} - 1\right) \pm dR \qquad (3)$$

where dR is the change due to recombination and is given by

$$dR = c(w_{14} f_1 f_4 - w_{23} f_2 f_3).$$

The sign of dR in (3) is negative for the coupling heterozygotes ($j = 1,4$) and positive for the repulsion heterozygotes ($j = 2,3$). If the coupling and repulsion heterozygotes are assumed to have the same selective advantage, then $dR = c \Delta w_{14}$. Substitution of (2) into (3) gives the general formula for response to truncation selection

$$df_j = \frac{i}{\sigma} f_j \left(\sum_k f_k v_{jk} - m\right) \pm c \Delta \left[1 + \frac{i}{\sigma} (v_{14} - m)\right]. \qquad (4)$$

Consider now the special case of two additive loci. Let them have genotypic effects, defined relative to the phenotypic

standard deviation, of magnitude

| AA | Aa | aa | BB | Bb | bb |
|---|---|---|---|---|---|
| $\alpha\sigma$ | $\alpha\sigma/2$ | 0 | $\beta\sigma$ | $\beta\sigma/2$ | 0 |

(5)

For the model (5) the population mean is $m = (p\alpha + q\beta)\sigma$. Letting $\mu = m/\sigma = p\alpha + q\beta$, the changes in gametic frequency from one cycle of selection are, from (4)

$$\left.\begin{aligned}
AB &: df_1 = \tfrac{i}{2} f_1 (\alpha + \beta - \mu) - c\Delta[1 + \tfrac{i}{2}(\alpha + \beta - 2\mu)] \\
Ab &: df_2 = \tfrac{i}{2} f_2 (\alpha - \mu) + c\Delta[1 + \tfrac{i}{2}(\alpha + \beta - 2\mu)] \\
aB &: df_3 = \tfrac{i}{2} f_3 (\beta - \mu) + c\Delta[1 + \tfrac{i}{2}(\alpha + \beta - 2\mu)] \\
ab &: df_4 = \tfrac{i}{2} f_4 (-\mu) - c\Delta[1 + \tfrac{i}{2}(\alpha + \beta - 2\mu)]
\end{aligned}\right\} \quad (6)$$

The changes in gene frequency are

$$\left.\begin{aligned}
A &: dp = df_1 + df_2 = \tfrac{i}{2}[\alpha p(1-p) + \beta\Delta] \\
B &: dq = df_1 + df_3 = \tfrac{i}{2}[\beta q(1-q) + \alpha\Delta] .
\end{aligned}\right\} \quad (7)$$

From (7) it can be seen that $i\alpha$ and $i\beta$ are the usual selective values of the alleles A and B respectively. Also equation (7) illustrates that $\Delta$ is the covariance of the allelic frequencies of A and B, so that the change in the frequency of A results from both the direct response from selection on A and a correlated response from selection on B. The disequilibrium, $\Delta$, can be seen to be a covariance by rearranging equation (1) into the form $\Delta = f_1 - pq$. Thus $\Delta$ equals the frequency of gametes containing both A and B less the product of their marginal frequencies and is therefore the covariance of a bivariate binomial distribution (Kendall, 1943, p.133). In terms

of the disequilibrium determinant, the additive genetic variance ($\sigma_A^2$) reduces to the same form as (7):

$$\sigma_A^2 = [\frac{\alpha^2}{2} p(1-p) + \frac{\beta^2}{2} q(1-q) + \alpha\beta\Delta]\sigma^2. \qquad (8)$$

If a population is initially in linkage equilibrium, then A. Robertson (personal communication) has shown that after one cycle of truncation selection on the individual phenotype with additive genes

$$\Delta_1 = \frac{1}{4}(ix - i^2)\,\alpha\beta\, p(1-p)\, q(1-q) \qquad (9)$$

where x is the truncation point in standard units of the phenotypic distribution, which is assumed to be normal. Felsenstein (1965) pointed out that the initial disequilibrium would be negative, but did not give its magnitude. Since $(ix - i^2) \leq 0$, equation (9) also shows that selection will generate negative disequilibrium and thus reduce the response if linkage is tight. However if gene effects are small the amount of disequilibrium actually generated will be very small with an additive model, and long term selection response is unlikely to be greatly affected by the degree of recombination in an infinitely large population.

If the initial build up of $\Delta$ from a population in equilibrium is calculated from the changes in gametic frequency (6), the prediction turns out to be (Nei, 1963)

$$\Delta_1 = -\frac{1}{4} i^2\, \alpha\beta p(1-p)q(1-q) \qquad (10)$$

$$= -dpdq$$

The discrepancy between equations (9) and (10) arises from the fact that for (10) the selective advantages of the gametes (equation 2) were

calculated by excluding terms in squared proportionate effects ($\alpha^2$, $\alpha\beta$, $\beta^2$), whereas these were included in computing (9). Equation (9) is therefore more precise, but for simplicity the selective advantages (2) will be used to compute changes in gametic frequency in the Monte Carlo study to follow.

In general, using the approximate selective values (2) and changes in gametic frequency (6), the change in disequilibrium is given by

$$\Delta_{t+1} = \Delta_t (1 - c)[1 + \tfrac{i}{2} (\alpha + \beta - 2\mu)] - dpdq \qquad (11)$$

where dp and dq are given by (7).

For a population which is initially in linkage equilibrium to remain exactly in equilibrium after selection, the selective advantages of the gametes ($w_j = \sum_k f_k w_{jk}$) must be multiplicative. If, before selection

$$\Delta = f_1 f_4 - f_2 f_3 = 0$$

and after selection

$$(f_1 + df_1)(f_4 + df_4) - (f_2 + df_2)(f_3 + df_3) = 0$$

then from (3), it follows that

$$w_1 w_4 = w_2 w_3. \qquad (12)$$

Model (12) will be used in the small population study for comparison with the additive model and will be discussed in more detail later.

Many loci.

Geiringer (1944) developed the theory of recombination between many loci, and Bennett (1954) used Geiringer's results to extend the formulae for linkage disequilibria to more than two loci. However it

turns out that with non-epistatic genes disequilibria among more than pairs of loci do not appear in any equations for the additive variance or changes in gene frequency.

For n loci each of two alternative alleles $A_j$, $a_j$, the second order disequilibria are defined as in equation (1). The disequilibrium between loci j and k, $\Delta_{jk}$, is given by

$$\Delta_{jk} = f(A_jA_k) - p_jp_k$$

where $f(A_jA_k)$ is the sum of the frequencies of all gametes containing $A_j$ and $A_k$, and $p_j$ and $p_k$ are the gene frequencies of $A_j$ and $A_k$ respectively. If $c_{jk}$ is the recombination fraction between these loci, at generation t

$$\Delta_{jk(t)} = (1 - c_{jk})\Delta_{jk(t-1)}$$

$$= (1 - c_{jk})^t \Delta_{jk(o)}.$$

If the loci are additive, with effects

| $A_jA_j$ | $A_ja_j$ | $a_ja_j$ |
|---|---|---|
| $\alpha_j\sigma$ | $\alpha_j\sigma/2$ | 0 |

formula (8) for the additive variance can be extended to give

$$\sigma_A^2 = [\frac{1}{2} \sum_{j=1}^{n} \alpha_j^2 p_j(1-p_j) + \sum_j \sum_k \alpha_j\alpha_k \Delta_{jk}]. \qquad (13)$$

Similarly, with truncation selection, it can be shown that the change in gene frequency in one generation is

$$dp_j = \frac{i}{2} [\alpha_j p_j (1-p_j) + \sum_{k\neq j} \alpha_k \Delta_{jk}]. \qquad (14)$$

Formulae of the type (13) and (14) can readily be extended to include multiple alleles and dominance. However with epistatic gene action the partition of the genotypic variance into additive, dominance and epistatic components is less straightforward, because the effects are difficult to partition orthogonally if there is linkage disequilibrium. These extensions will not be pursued, for the small population study is concerned entirely with additive loci, for which the results of this section have been developed as a background.

## 3. BASIC THEORY FOR SMALL POPULATIONS

### One locus

Robertson's (1960) theory of limits to artificial selection in small populations was developed from some results of Kimura (1957). The concepts underlying their work were the distribution of gene frequencies and the chance of fixation of a gene. The gene frequency distribution can be regarded as either the distribution of the frequencies of loci of the same effect and magnitude in one population, or of an individual gene in many populations. Similarly, the chance of fixation of a gene can be considered either as the proportion of genes of the same kind fixed in a line, or as the proportion of replicate lines in which this gene is fixed, after a selection limit has been reached. The case where no further selection response can be made but not all the genes have become fixed due to heterozygote superiority or opposing natural selection will not be discussed.

Kimura (1957) used a continuous model to describe the change in the distribution of gene frequency, $\phi(p, t)$, at time t by means of the diffusion equation

$$\frac{\partial \phi}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial p^2} [V(dp)\phi] - \frac{\partial}{\partial p} [M(dp)\phi] \qquad (15)$$

where M(dp) and V(dp) are respectively the mean and variance of the change in gene frequency per generation. If the effective population size is N then $V(dp) = p(1-p)/2N$ and for additive genes with selective values $i\alpha$, $M(dp) = \frac{i\alpha}{2} p(1-p)$. Substituting for V(dp) and M(dp) in (15) gives after rearrangement

$$\frac{\partial \phi}{\partial (t/N)} = \frac{1}{4} \frac{\partial^2}{\partial p^2} [p(1-p)\phi] - \frac{Ni\alpha}{2} \frac{\partial}{\partial p} [p(1-p)\phi]. \qquad (16)$$

Thus for a given value of gene frequency, p, the selection process can be described by the parameter $Ni\alpha$ on a time scale of t/N (Robertson, 1960). The chance of fixation, $u(p_o)$, of a gene with initial frequency $p_o$ was given by Kimura (1957), and for the additive model

$$u(p_o) = \frac{1 - e^{-2Ni\alpha p_o}}{1 - e^{-2Ni\alpha}} \qquad (17)$$

and is shown in Figure 1.

Two important assumptions are made in the diffusion approximation : firstly that the population size is sufficiently large that the distribution of gene frequencies can be considered continuous, whereas in fact only 2N+1 discrete values of gene frequency are possible; and secondly that selective values are small, so that terms in $(i\alpha)^2$ can be ignored relative to $i\alpha$ and 1/N. Ewens (1963) investigated the fit of formula (17) from the diffusion equation with the chance of fixation computed by matrix iteration for the discrete model with N=6 and $0 \leqslant i\alpha \leqslant 0.2$ and found good agreement. However, in order to reduce computing time in the Monte Carlo study it was necessary to use selective values as large as $i\alpha = 1.0$ and to extrapolate from small populations (N = 8, 16) to those of larger size, so that further checks on Kimura's (1957) formula for the chance of fixation (17) are given in Table 1. The values of $u(p_o)$ for N = 8, 16 and 32 were calculated by repeated iteration of a matrix of transition probabilities of gene frequencies onto a vector of the

TABLE 1 The chance of fixation of a gene with selective value $1+\alpha$ calculated by matrix iteration for different population sizes (N) and by diffusion approximation ($N = \infty$).

| Nα | N | Initial frequency .05 | .1 | .3 | .5 | .7 |
|---|---|---|---|---|---|---|
| 32 | ∞ | .9592 | .9983 | 1.0000 | 1.0000 | 1.0000 |
| 32 | 32 | .9412 | .9967 | 1.0000 | 1.0000 | 1.0000 |
| 16 | ∞ | .7981 | .9592 | .9999 | 1.0000 | 1.0000 |
| 16 | 32 | .7766 | .9507 | .9999 | 1.0000 | 1.0000 |
| 16 | 16 | .7591 | .9434 | .9999 | 1.0000 | 1.0000 |
| 8 | ∞ | .5507 | .7981 | .9918 | .9997 | 1.0000 |
| 8 | 32 | .5392 | .7883 | .9909 | .9996 | 1.0000 |
| 8 | 16 | .5291 | .7797 | .9901 | .9996 | 1.0000 |
| 8 | 8 | .5124 | .7653 | .9890 | .9996 | 1.0000 |
| 4 | ∞ | .3298 | .5509 | .9096 | .9820 | .9966 |
| 4 | 32 | .3260 | .5462 | .9076 | .9817 | .9966 |
| 4 | 16 | .3226 | .5419 | .9059 | .9814 | .9966 |
| 4 | 8 | .3164 | .5343 | .9031 | .9811 | .9967 |
| 2 | ∞ | .1847 | .3358 | .7118 | .8808 | .9567 |
| 2 | 32 | .1838 | .3346 | .7108 | .8805 | .9567 |
| 2 | 16 | .1830 | .3334 | .7099 | .8802 | .9568 |
| 2 | 8 | .1815 | .3312 | .7083 | .8799 | .9570 |
| 1 | ∞ | .1101 | .2096 | .5218 | .7311 | .8713 |
| 1 | 32 | .1099 | .2094 | .5216 | .7310 | .8714 |
| 1 | 16 | .1098 | .2092 | .5213 | .7309 | .8714 |
| 1 | 8 | .1095 | .2087 | .5209 | .7308 | .8716 |

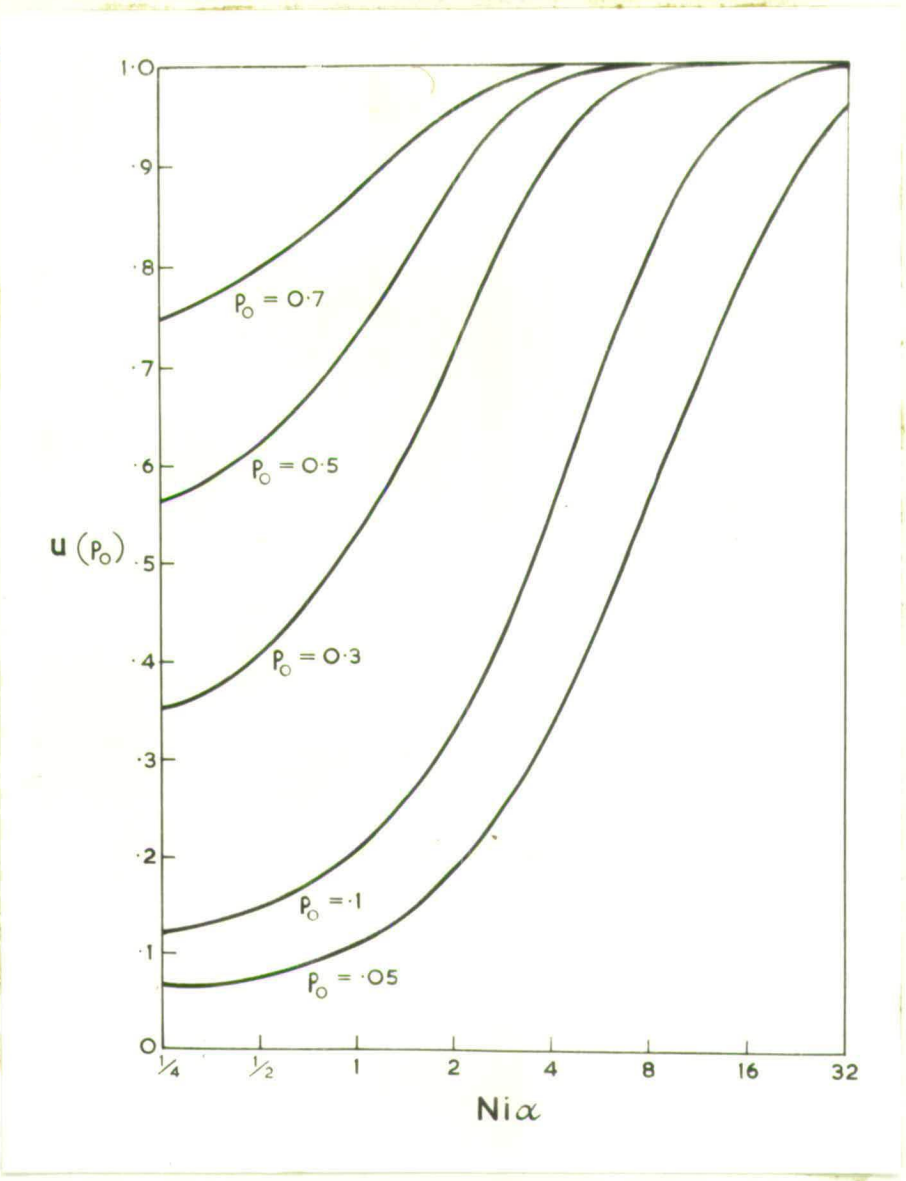


FIGURE 1 The chance of fixation of a single additive gene.

distribution of gene frequencies. Agreement of the matrix and diffusion equation results is generally very good, perhaps remarkably so since some of the selective values used are so large. The poorest agreement is found with low initial frequencies; but for the parameters studied, notably $i\alpha \leq 1.0$, $p_o \geq 0.05$, $N \geq 8$, the response $(u(p_o) - p_o)$ predicted from matrix iteration never differs by more than 8% from that predicted by the continuous model, which can be regarded as the case of population size becoming infinitely large.

The diffusion equation for linked loci

With linked loci it is convenient to study the distribution of gametic frequencies rather than gene frequencies. In addition to selection and drift a third force is acting on the distribution, that of recombination. The changes in the distribution of gametic frequencies can be described by a continuous model, the multi-dimensional diffusion equation (e.g. Kimura, 1955) of the general form

$$\frac{\partial \phi}{\partial t} = \frac{1}{2} \sum_{j=1}^{h} \frac{\partial^2}{\partial f_j^2} [V(df_j)\phi] + \sum_j \sum_k \frac{\partial^2}{\partial f_j \partial f_k} [Cov(df_j, df_k)\phi]$$

$$- \sum_{j=1}^{h} \frac{\partial}{\partial f_j} [M(df_j)\phi] \qquad (18)$$

where $\phi(f_1, \ldots, f_h, t)$ is the distribution of gametic frequencies, $f_j$, at time t. The dimension of the equation is h, and is equal to the number of degrees of freedom amongst the gametic frequencies. For n loci each of two alleles $h = 2^n - 1$. From the multinomial distribution,

the variance of the change in gametic frequency is given by $V(df_j) = [f_j(1 - f_j)]/2N$, and the covariance of changes by $Cov\ (df_j, df_k) = -(f_j f_k)/2N$. Directional changes in gametic frequency, $M(df_j)$, are given by (3), or by (4) for truncation selection. For the diffusion equation to hold, both the recombination fraction (c) and selective values $(w_{jk} - 1)$ should not be of greater order than $1/N$, so that terms in their product can be ignored. Thus, after rearrangement, the diffusion equation for truncation selection with two loci turns out to be

$$\frac{\partial \phi}{\partial (t/N)} = \frac{1}{4} \sum_{j=1}^{3} \frac{\partial^2}{\partial f_j^2} [f_j(1-f_j)\phi] - \frac{1}{2} \sum_{j}\sum_{k} \frac{\partial^2}{\partial f_j \partial f_k} [f_j f_k \phi]$$

$$- \frac{Ni}{\sigma} \sum_{j=1}^{3} \frac{\partial}{\partial f_j} [(\sum_{k=1} f_k v_{jk} - m)\phi] \mp Nc \sum_{j=1}^{3} \frac{\partial}{\partial f_j} [\Delta \phi] \qquad (19)$$

where $f_4$ must be formally replaced by $1-f_1-f_2-f_3$ and $\Delta$ by $f_1(1-f_1-f_2-f_3)-f_2 f_3$. Clearly, if the terms $\sum_{k=1}^{4} f_k v_{jk}$ and m in (19) are written out in full, the $v_{jk}$ can be taken out of the differentials as constants. Thus, similarly, to the one dimensional equation (16), on a time scale of t/N and for a given set of frequencies the process can be described by the parameters $\frac{Ni}{\sigma} v_{11}, \frac{Ni}{\sigma} v_{12}, \ldots\ldots, \frac{Ni}{\sigma} v_{44}$ and Nc. For the special case of two additive loci the parameters reduce to only $Ni\alpha$, $Ni\beta$ and Nc where $\alpha$ and $\beta$ are defined by (5).

The chance of fixation of an additive gene, A, acting alone in a population was seen to be a function of $p_o$, its initial frequency, and $Ni\alpha$. It has now been shown that if another additive gene, B, is also segregating a further four parameters can affect the chance of fixation of A. These are $Ni\beta$ and Nc, measures of the effect of B and

the tightness of linkage of A to B, $q_o$, the initial frequency of B, and $\Delta_o$, the initial linkage disequilibrium. Much of the work to be described will be concerned with estimating how these extra parameters influence the chance of fixation of A. However it will usually be assumed that the population is initially in linkage equilibrium ($\Delta_o = 0$).

No algebraic solution of (19) has been obtained, even for the additive model. Numerical solution of the differential equation could have been attempted but was likely to involve excessive computing time and storage. Although simulation of the process of selection in small populations by means of transition probability matrix iteration has been used successfully for one locus (Allan and Robertson, 1964), the method is not practicable for two or more loci. The two locus model requires a square transition matrix of dimension $\binom{2N+3}{3}$ which is, for example, 165 with a population size N = 4. It was therefore decided to study the system by means of Monte Carlo simulation.

The breakdown of linkage disequilibrium

In this section formulae are derived for the rate of breakdown of linkage disequilibrium in unselected small populations and the results used to calculate the response to artificial selection when some simplifying assumptions are made.

In a population of effective size N in which no selection is practised the expected disequilibrium after one generation is given by

$$\Delta_1 = E[(f_1 + df_1)(f_4 + df_4) - (f_2 + df_2)(f_3 + df_3)].$$

Using the multinomial distribution, it can be shown that

$$\Delta_1 = (1 - c)(1 - 1/2N)\Delta_o$$

If c and 1/2N are small such that c/2N can be ignored relative to c and 1/2N then

$$\Delta_1 = (1 - c - 1/2N)\Delta_o$$

$$\text{and} \quad \Delta_t = (1 - c - 1/2N)^t \Delta_o$$

$$\sim \Delta_o \; e^{-(2Nc + 1)t/2N}. \qquad (20)$$

Thus on a time scale of $t/N$, $\Delta_t/\Delta_o$ is a function of $(2Nc + 1)$. The half life of the breakdown of $\Delta$, at which $\Delta_t/\Delta_o = 0.5$, occurs when

$$t = \frac{1.4N}{(2Nc + 1)} \quad \text{generations} \qquad (21)$$

In Figure 2 the function $1/(2Nc + 1)$ is plotted against Nc, with Nc on a log scale. The slope has a maximum at $2Nc = 1$, so that the rate of breakdown of disequilibrium is most sensitive to multiplicative changes in recombination fraction at values of c near 1/2N. The function $1/(2Nc + 1)$ also appears in formulae for the probability of recombination in small populations inbred to fixation (Wright, 1933; Kimura, 1963).

If selection is very weak and gene effects are small, it can be assumed that the mean gene frequency and hence the variance and the disequilibrium change very little as a result of selection. The additive variance for one locus therefore declines as

$$\sigma^2_{A(t)} = \sigma^2_{A(o)} e^{-t/2N}$$

and $\Delta$ declines as in equation (20). Therefore for a multi-locus additive model, for which the variance in the first generation is given

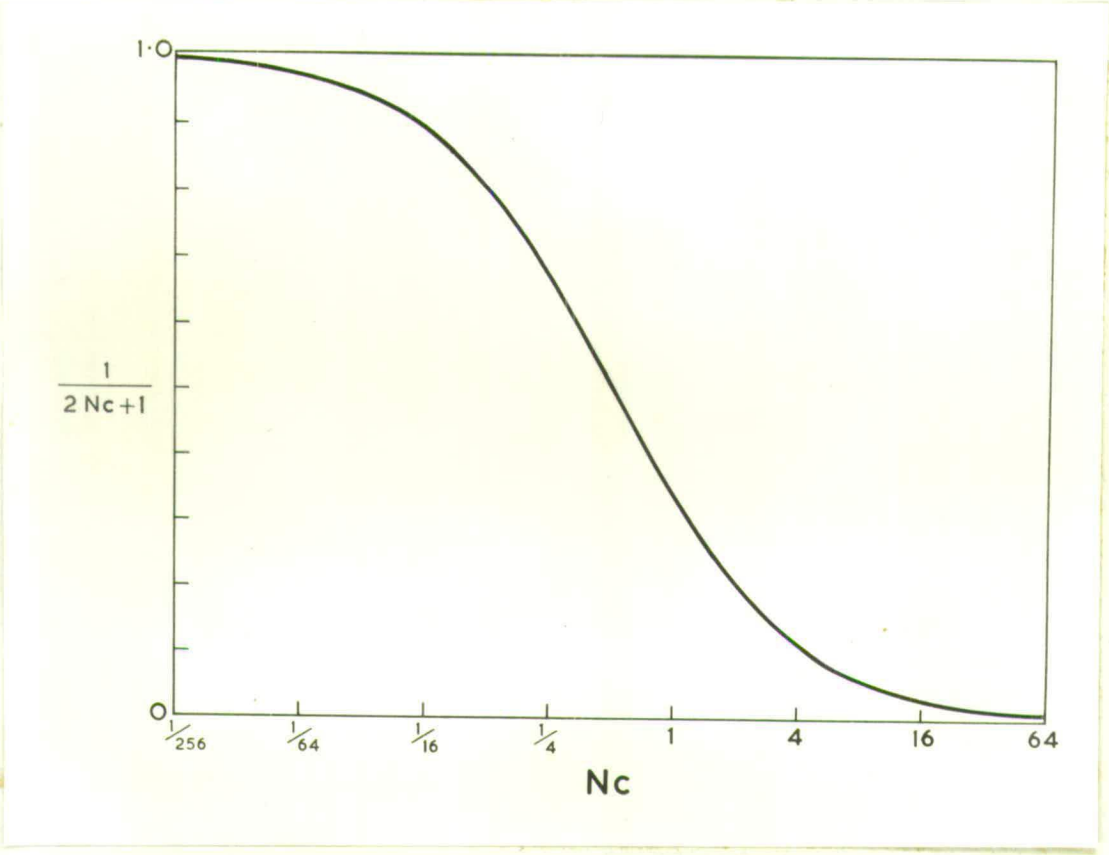


FIGURE 2 Graph of 1/(2Nc + 1) against Nc.

by (13), the variance at generation t is

$$\sigma^2_{A(t)} = [\tfrac{1}{2} \sum_j \alpha_j^2 p_j(1 - p_j)\, e^{-t/2N} + \sum_j \sum_k \alpha_j \alpha_k \Delta_{jk}\, e^{-(2Nc_{jk} + 1)t/2N}]\, \sigma^2,$$

and the total selection advance is given by

$$\int_0^\infty \frac{i\sigma^2_{A(t)}}{\sigma}\, dt = Ni\sigma [\sum_j \alpha_j^2 p_j(1 - p_j) + 2 \sum_j \sum_k \alpha_j \alpha_k \Delta_{jk}/(2Nc_{jk} + 1)]. \quad (22)$$

Similarly, the total change in the frequency of a gene can be expressed as

$$u(p_o) - p_o = Ni[\alpha_j p_j(1 - p_j) + \sum_{k \neq j} \alpha_k \Delta_{jk}/(2Nc_{jk} + 1)] \quad (23)$$

The above equations, (22) and (23), were derived using the approach of Robertson (1960) who gave similar formulae for a single gene and showed that with weak selection the total response would be 2N times that in the first generation. Clearly, from (22) and (23) it can be seen that if there is negative disequilibrium initially then the total advance will exceed 2N times that in the first generation and if there is initial positive disequilibrium less than 2N times the first generation's response will be made, even with very weak selection. The assumptions in the above derivations are very strong, and usually the total advance will be less than 2N times that in the first generation, even with one locus, the discrepancy becoming greater the larger $Ni\alpha$ (Robertson, 1960). From this simple model however, one important conclusion can be salvaged. If there is initial linkage equilibrium and if the parameters $Ni\alpha_j$ are very small, then the population remains in equilibrium and the selection limit is not influenced by the tightness of linkage.

## 4. SIMULATION PROCEDURE

The simulation process used in this Monte Carlo study differed from that of other workers (Fraser, 1957a; Martin and Cockerham, 1960; Gill, 1963, 1965; Qureshi, 1963). They simulated gametes on the computer and paired these to form individual genotypes. The genotypic value of each individual was specified by the mode of gene action used, and the phenotypic value computed by adding to the genotypic value a random normal deviate as environmental error. Selection of parents for the next generation was based on these phenotypic values. New gametes were formed from the parents in which crossing over between adjacent loci occurred with probability specified by the recombination fractions.

In the procedure used here the 2N gametes formed each generation were never paired into genotypes. The expected frequencies of the gametes ($f_j$) in the next generation were calculated algebraically, and the calculation included both selection and recombination. Thus for truncation selection with an additive model of two loci, the new expected gametic frequencies were computed by formula (6). In each run the 2N gametes in the next generation were obtained by sampling from a multinomial distribution with parameters $f_j$ by means of generating 2N uniform pseudo-random numbers, x, and comparing each with the gametic frequencies. If

$0 < x \leq f_1$ then a gamete AB was generated, or if

$f_1 < x \leq f_1 + f_2$ then a gamete Ab was generated, and so on.

In the computer programme each of the parameters N, $i\alpha$, $i\beta$, c and the initial frequencies could be altered. From the initial frequencies for any run selection was practised first (e.g. by formula 6), before random gametes were formed in the manner described above.

Each iteration was continued to fixation or for 6.25 N generations, whichever occurred first. After 6.25 N generations of selection for one locus at least 99.9% of the total response can be expected to be made if $Ni\alpha \geqslant 4$, 98.5% if $Ni\alpha = 2$ or 96.6% if $Ni\alpha = 1$. The average gene frequency after 6.25 N generations was therefore taken as the limit whether complete fixation in all lines had, or had not, taken place. Usually 400 replicates were run for each set of parameters. At fixation, the proportion of lines in which the favourable gene is fixed is binomially distributed, so that with 400 replicates the chance of fixation of the favourable gene, $u(p_o)$, has a standard error of $0.05 \sqrt{u(p_o)[1 - u(p_o)]}$.

An I.C.T. Atlas Computer was used for the simulation. It is a fast machine (by present standards), such that for a population size of 8, 400 replicates each of 50 generations required about 12 seconds of computing time, dependent on the rate of fixation. Doubling the population size increased the computing time by a factor of almost four.

The formulae used to calculate the changes in gametic frequency were derived for infinite populations so they are not precise for the small population sizes of 8 or 16 studied. In particular, an assumption of the infinite model is that each genotypic frequency is exactly the product of its constituent gametic frequencies. However the objective of this study was not to obtain results applicable only

to one population size but to be able to extrapolate to populations usually larger than those simulated. It was therefore thought advisable to adopt general formulae for selection response within populations and consider sampling only of gametes for the next generation. Furthermore, considerable savings in computing time could be made by using the algebraic approximations to selection response and expected amount of recombination.

## 5. THE SELECTION LIMIT

### The chance of fixation of a linked gene

In the first part of the Monte Carlo study all possible combinations of several values of the parameters $p_o$, $q_o$, $Ni\alpha$, $Ni\beta$ and Nc were run for two additive genes initially in linkage equilibrium. These were:

Initial frequencies, $p_o$, $q_o$ = 0.05, 0.1, 0.3, 0.5, 0.7;

$Ni\alpha$. $Ni\beta$ = 2, 4, 8, 16;

Nc = 1, 1/4, 0.

To avoid selective values $i\alpha$, $i\beta$ greater than one, combinations in which $Ni\alpha$ or $Ni\beta$ = 16 were run at a population size N = 16. All other combinations were run with N = 8 to reduce computing time. For both population sizes 400 replicates were used. The chance of fixation of the gene with initial frequency $p_o$ and effect $\alpha$ is shown for the above combinations of parameters in Figures 3-7. In addition, the chance of fixation for each value of $p_o$ and $Ni\alpha$ is given from the matrix iteration for one locus (as Table 1, but taken for only 6.25 N generations) with N = 8 for $Ni\alpha$ = 2, 4, 8 and N = 16 for $Ni\alpha$ = 16. These results are labelled Nc = $\infty$, for the response of a gene acting alone can also be viewed as the response of that gene when segregating independently of other genes in the population, implying free recombination in a very large population. The matrix results do not correspond to free recombination (i.e. c = 0.5) in a population of relatively small size, when the maximum value of Nc is N/2, or 4 for a population size of 8.

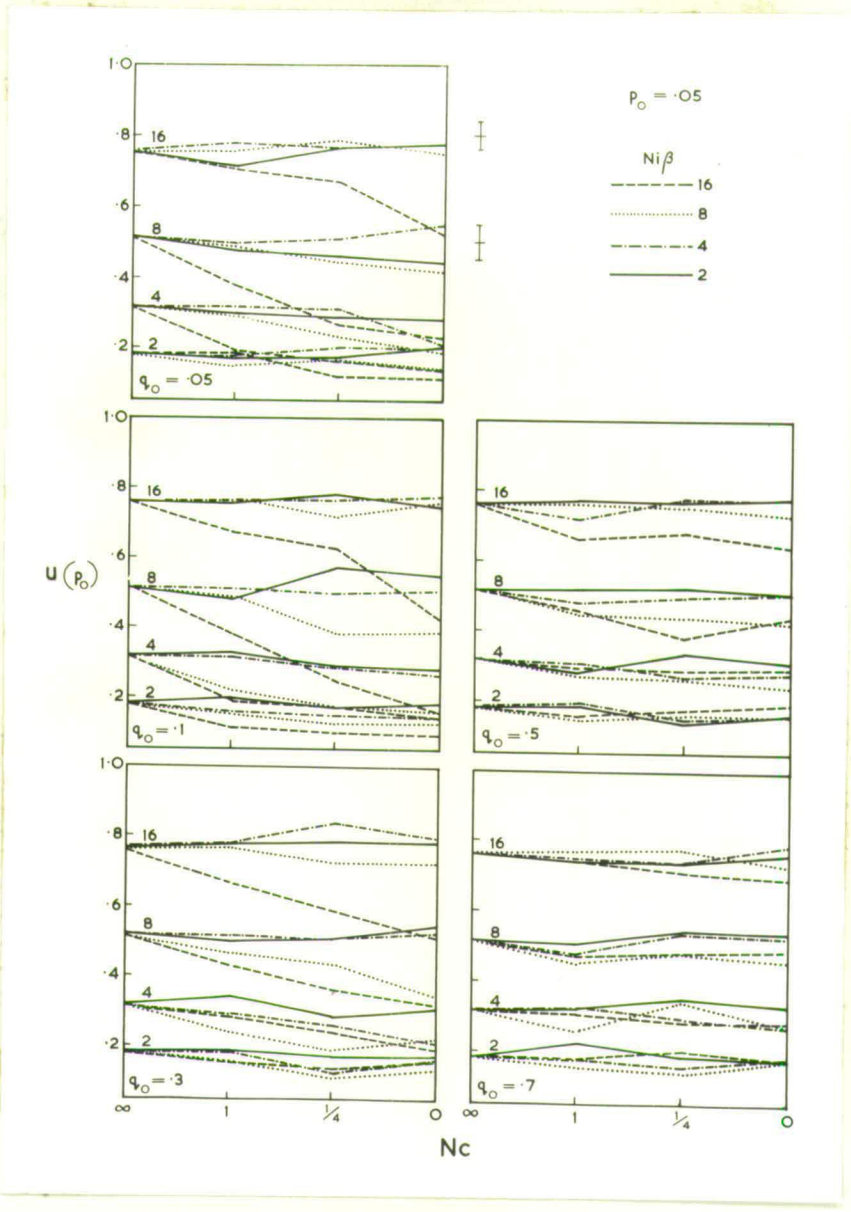


FIGURE 3 The influence of a linked additive gene on the chance of fixation of an additive gene with initial frequency 0.05 for various values of Ni$\alpha$. Typical ranges of length four standard deviations are also shown.

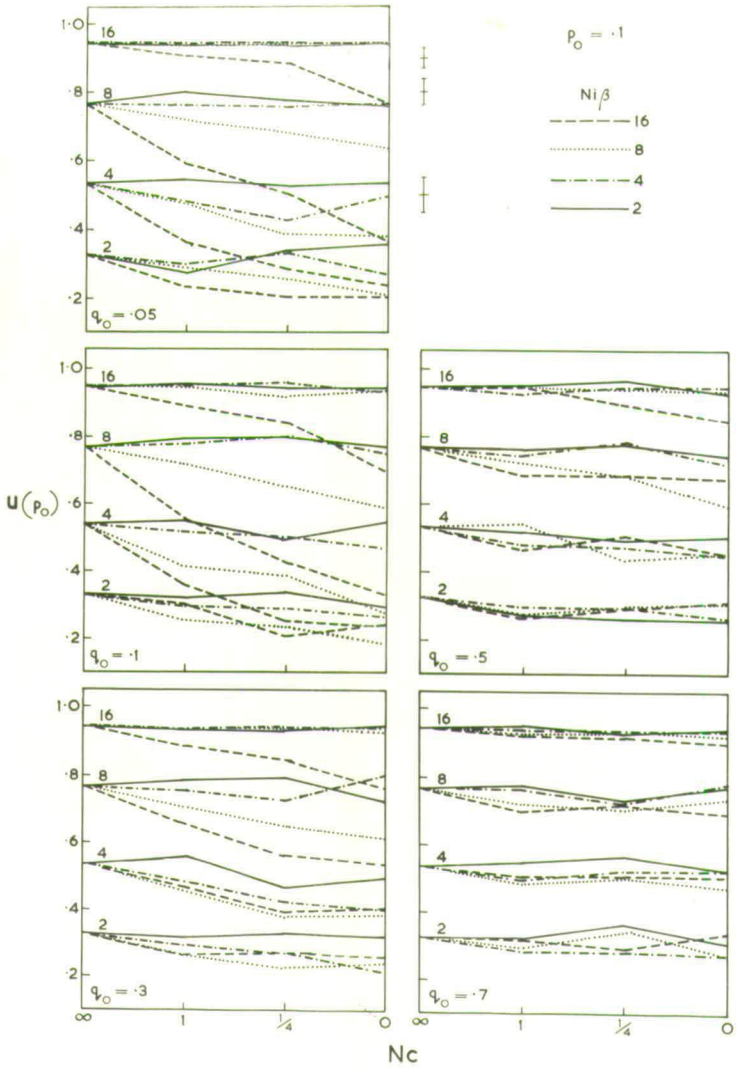


FIGURE 4 The influence of a linked additive gene on the chance of fixation of an additive gene with initial frequency 0.1 for various values of Nia. Typical ranges of length four standard deviations are also shown.

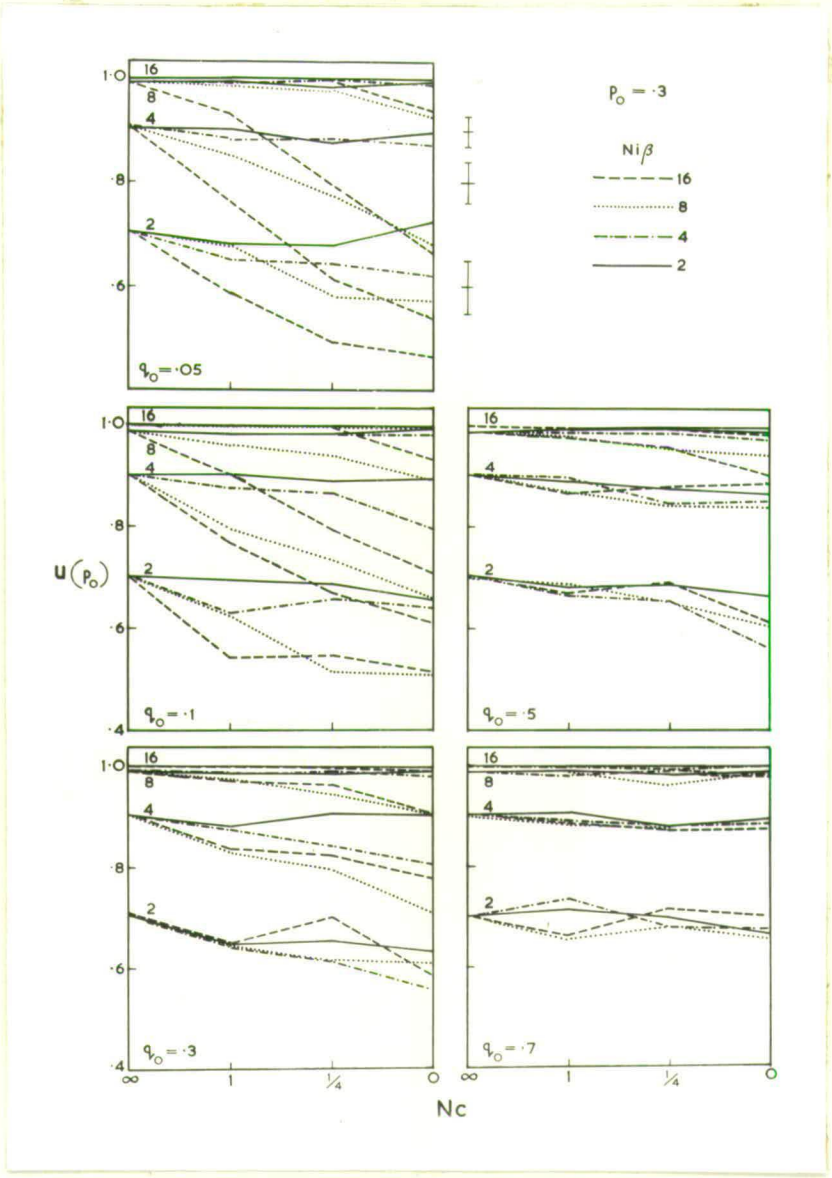


FIGURE 5 The influence of a linked additive gene on the chance of fixation of an additive gene with initial frequency 0.3 for various values of Nia. Typical ranges of length four standard deviations are also shown.

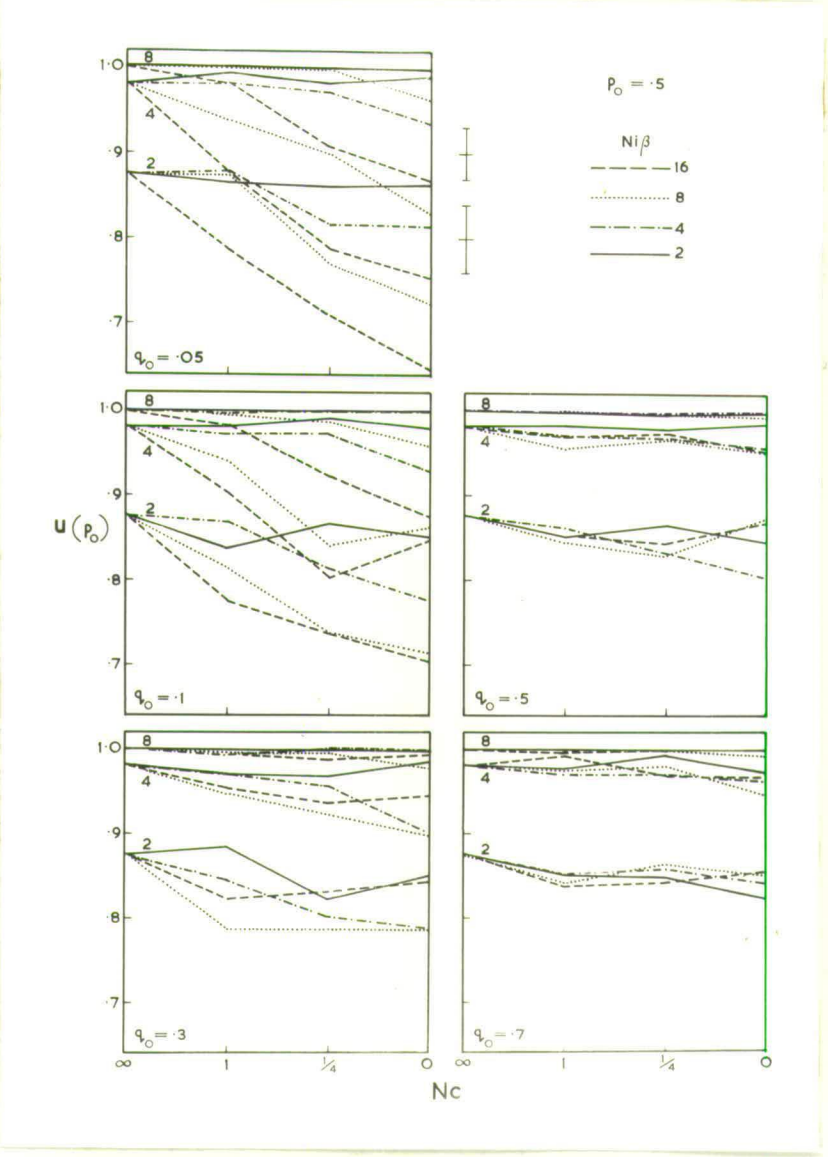


FIGURE 6 The influence of a linked additive gene on the chance of fixation of an additive gene with initial frequency 0.5 for various values of Ni$\alpha$. Typical ranges of length four standard deviations are also shown.

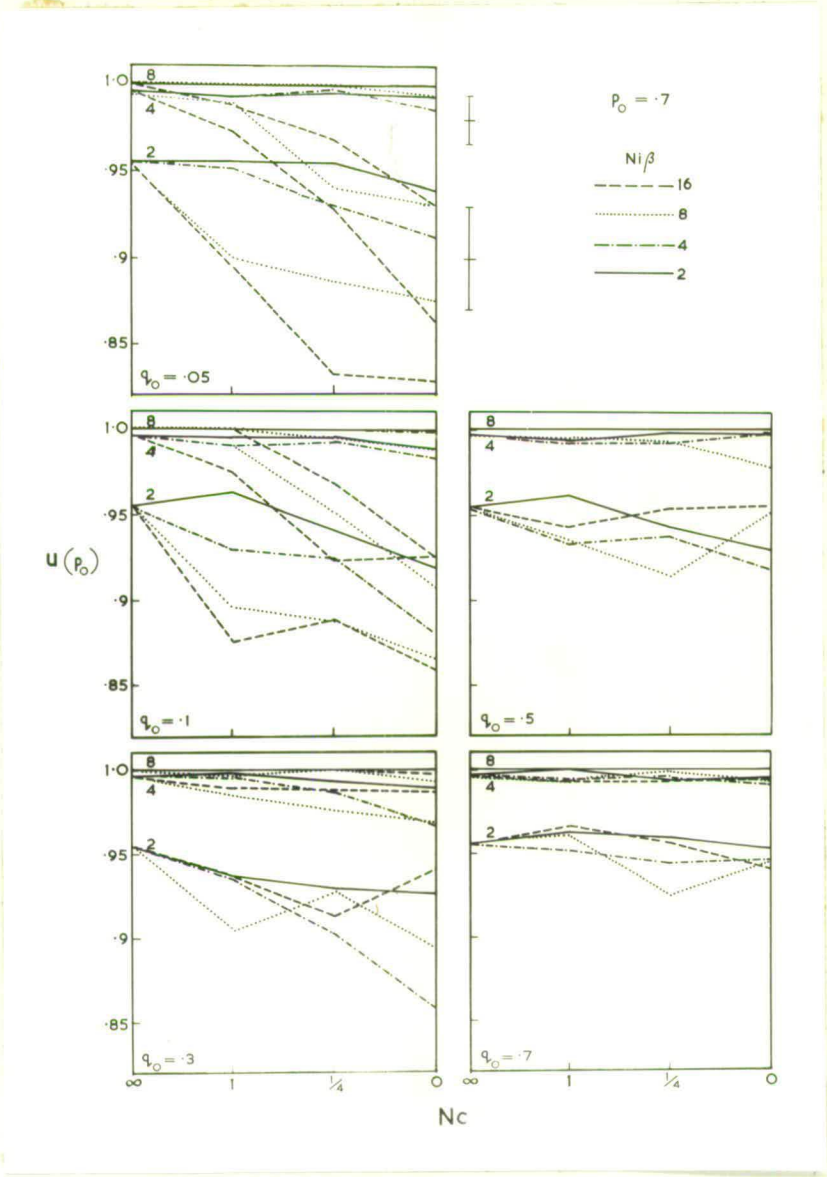


FIGURE 7 The influence of a linked additive gene on the chance of fixation of an additive gene with initial frequency 0.7 for various values of Nia. Typical ranges of length four standard deviations are also shown.

The simulation results for $N_i\alpha = 2$, 4 and 8 which were run at $N = 16$ (i.e. $N_i\beta = 16$) have been adjusted to make them comparable with those run with $N = 8$, since there are small differences in chance of fixation for constant $N_i\alpha$, but varying N (Table 1). The results for $N = 16$ were multiplied by the factor

$$\frac{u(p_o,\ 8) - p_o}{u(p_o,\ 16) - p_o} \qquad (24)$$

which standardises responses, where $u(p_o, N)$ is the chance of fixation of the gene segregating independently in a population of size N, computed by matrix iteration. Although the correction was arbitrarily chosen, it usually makes small changes relative to the standard error of the Monte Carlo estimates and does not affect any of the conclusions to be drawn from the results. The main advantage in making some transformation of the form (24) is that the same value of $u(p_o)$ can be plotted in Figures 3-7 for all N with $N_c = \infty$ .

The data in Figures 3-7 for two linked additive genes, A with effect $\alpha$ and initial frequency $p_o$, and B with effect $\beta$ and initial frequency $q_o$, which are initially in linkage equilibrium show that:

(a) The chance of fixation, $u(p_o)$, of A may be greatly reduced if A is tightly linked to B, relative to A's chance of fixation if segregating independently. It is clear from the standard errors of the estimates of $u(p_o)$ shown in Figures 3-7 that very highly significant reductions in chance of fixation occur with many sets of parameters.

(b) The largest reductions in $u(p_o)$ are found when B has a low frequency.

(c) At least for low frequencies of the interfering gene B, and for sizes of $N_i\beta$ studied, the greater the effect of B, the greater the reduction in $u(p_o)$, the chance of fixation of the affected gene A. Also there is apparently no change in $u(p_o)$ if $N_i\beta \leq N_i\alpha/2$.

(d) The chance of fixation of the affected gene can be reduced for any value of its initial frequency except, perhaps, when the gene is almost certain to be fixed ($u(p_o) \sim 1$) when segregating independently.

(e) The tighter the linkage between the two genes, then the greater is the reduction in $u(p_o)$.

These observations clearly need further examination, so the data of Figures 3-7, together with additional results simulated for some particular examples of parameters, will be investigated in greater detail. The degree of recombination will be considered first.

It can be seen in Figures 3-7 that for a wide range of parameter sets $p_o$, $q_o$, $N_i\alpha$ and $N_i\beta$, the decline of $u(p_o)$ with tighter linkage is approximately linear for the spacings of Nc used in these graphs. The transformation used to plot these Nc values was the function $1/(2N_c + 1)$ which was shown in Figure 2 and gives values of $1/(2N_c + 1) = 0, 1/3, 2/3$ and 1 for $N_c = \infty, 1, 1/4$ and 0, respectively. This transformation was chosen partly for convenience, as it reduces values of Nc ranging from zero to infinity to a scale ranging only from zero to one, but mainly because the transformed variate is a measure of the rate of breakdown of linkage disequilibrium. It was shown earlier that on a time scale of t/N generations, $1/(2N_c + 1)$ is proportional to the half life of the breakdown of the disequilibrium determinant, $\Delta$, in small populations (equation 21).

The regression of chance of fixation on $1/(2Nc + 1)$ was computed for each set of parameters $p_o$, $q_o$, $Ni\alpha$ and $Ni\beta$ using data unadjusted by formula (24). As the variance of an estimate of $u(p_o)$ from some Monte Carlo simulation (i) depends on $u(p_o)$, squared deviations from the regression were weighted by $1/\sigma_i^2$, where $\sigma_i^2$ is the variance of the estimate. The sampling variance of each Monte Carlo estimate was calculated from the observed chance of fixation, so that repeated re-estimation of the variance was not required for each possible regression line. The chance of fixation for Nc was first compared with that for $Nc = \infty$, then regression lines were forced through $u(p_o)$ for $Nc = \infty$. The latter was calculated by matrix iteration and, of course, has no sampling variance. No analysis was performed where $u(p_o) > .99$ for $Nc = \infty$, for at such high frequencies errors in estimation of the variance become more serious, and the estimates of $u(p_o)$ and its variance more highly correlated. In addition to the data shown in Figures 3-7, computer runs were also made with N = 8 for the same parameter combinations shown in Figures 3-7, but with $Ni\alpha$, $Ni\beta < 16$, and Nc = 4 (free recombination with N = 8) and Nc = 1/4. The data for the regression analysis thus comprised 280 parameter sets, of which 100 had three estimated points: $1/(2Nc + 1) = 1/3, 2/3, 1$ and 180 had five estimated points: $1/(2Nc + 1) = 1/9, 1/3, 2/3, 8/9, 1$. Pooled results are given in Table 2, where sums of squares are of the form $\Sigma(d_i^2/\sigma_i^2)$ where $d_i$ are deviations from $Nc = \infty$ or regression and $\sigma_i^2$ the sampling variance of the estimate of $u(p_o)$.

TABLE 2 Pooled analysis of linear regressions of $u(p_o)$ against $1/(2Nc + 1)$.

| Source | Sum of squares | d.f. |
|---|---|---|
| Total (Deviations from $u(p_o)$ for $Nc = \infty$ ) | 15893 | 1200 |
| Linear regressions | 14584 | 280 |
| Residual from fitting linear regressions. | 1309 | 920 |

With the large number of replications used, the Monte Carlo estimates of $u(p_o)$ approach a normal distribution, so that under a null hypothesis of no effects $\frac{d_i^2}{\sigma_i^2}$ is distributed as chi-square with one degree of freedom. The sum of n such independent $\chi_1^2$ is distributed as $\chi_n^2$ and has expectation n. In Table 2 the total, and both its components, linear regressions and residual from fitting linear regression are significantly different from the appropriate $\chi^2$ ($P < .0001$ in each case). However, although the residual variance is significant, it contributes a very small proportion of the variability in this data. If binomial sampling variance is deducted from the sum of squares for the total and linear regressions, then the linear regressions remove 14304/14693 or 97.4% of the remaining variability for these parameter sets. The individual regression analyses most often show significant non linear regressions when the chance of fixation is very close to one, in which case the reduction for $Nc = 0$ exceeds that expected from intermediate Nc values, and when one gene is much larger than the other, $Ni\beta \gg Ni\alpha$, in which case the reduction for $Nc = 0$ is less than that expected from intermediate Nc values.

These exceptions to the general trend of a linear decline of $u(p_o)$ with increase in $1/(2Nc + 1)$ will be considered later.

These regressions also provide good evidence that tight linkage always reduces response with additive genes initially in equilibrium. In the 280 lines fitted, there was a significant ($P < .05$) linear regression in 176 cases. In all but one of these 176, the regression showed a reduction in $u(p_o)$ as linkage became tighter.

In Figures 3-7 it was shown that the influence of the affecting gene, B, on the chance of fixation of the gene A is highly dependent on the initial frequency and effect of B. The influence of the size and initial frequency, $q_o$, of B was studied in greater detail for a gene A with initial frequency $p_o = 0.3$ and with $Ni\alpha = 4$. Further runs with 400 replicates beyond those shown in Figures 3-7 were made for $q_o = 0.025$, 0.075, 0.2, 0.4 and 0.6, and also results were simulated for $Ni\beta = 32$ with $Nc = 0$, using a population size of 32. In Figure 8 the chance of fixation, $u(p_o)$, is plotted against $q_o$ for $Nc = 0$ and $Ni\beta = 2$, 4, 8, 16 and 32. The values of $u(p_o)$ from matrix iteration for one locus segregating are given in Figure 8 both for all $Ni\beta$ when $q_o = 0$ or $q_o = 1$ and also for a neutral gene, $Ni\beta = 0$, for all $q_o$. In neither case does selection change the frequency of the interfering gene, B, so that no influence on A's chance of fixation can be expected. It can be seen in Figure 8 that as $Ni\beta$ is increased, then, at least up to $Ni\beta = 16$, the maximum reduction in $u(p_o)$ is also increased. Further, the larger $Ni\beta$, the lower the initial frequency $q_o$ at which the maximum reduction takes place, such that for the larger $Ni\beta$ values it seems that the maximum reduction occurs where $Ni\beta q_o = 0.8$ approximately.

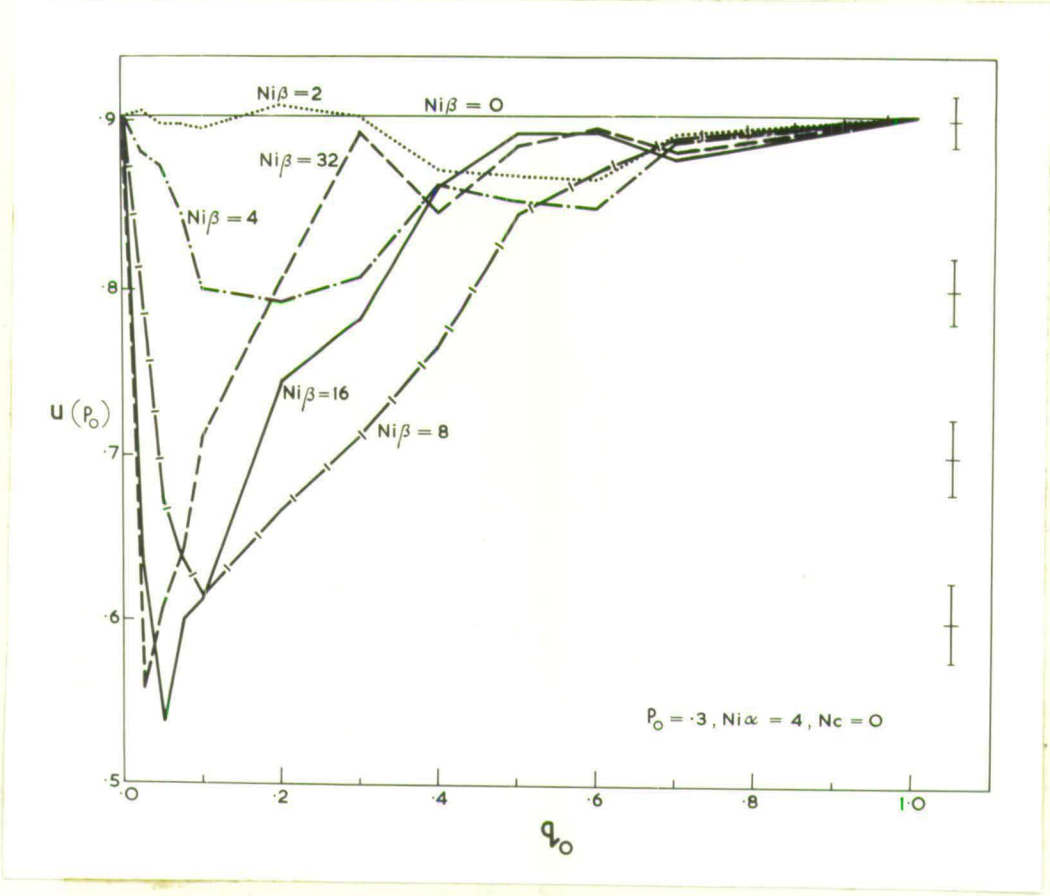


FIGURE 8 The relationship between the chance of fixation of an additive gene and the effect and initial frequency of another completely linked additive gene. Typical ranges of length two standard deviations are also shown.

However if B has an initial frequency higher than that causing the greatest reduction in $u(p_o)$ for a particular value of $Ni\beta$, then B may influence $u(p_o)$ less than does an interfering gene of smaller effect. For example, with $Ni\alpha = 4$, $p_o = 0.3$ and $Nc = 0$, the estimates of $u(p_o)$ are 0.675 for $Ni\beta = 8$ and 0.537 for $Ni\beta = 16$ with $q_o = 0.05$, but the estimates are 0.766 for $Ni\beta = 8$ and 0.863 for $Ni\beta = 16$ with $q_o = 0.4$. On the other hand, Figure 8 indicates that if B has a frequency lower than that causing the greatest reduction in $u(p_o)$ for some $Ni\beta$, then B influences $u(p_o)$ more than any gene of smaller effect and the same frequency.

Results for different degrees of linkage in the example of Figure 8 in which $p_o = 0.3$ and $Ni\alpha = 4$ are shown in Figure 9 for $Ni\beta = 2$ and $Ni\beta = 4$, in Figure 10 for $Ni\beta = 8$ and in Figure 11 for $Ni\beta = 16$. In these graphs chances of fixation from Monte Carlo simulation for $Nc = 1$ and $1/4$ are compared with those for $Nc = 0$ given in Figure 8 and with $Nc = \infty$ from one locus iteration. Figures 10 and 11, with the higher $Ni\beta$ values, indicate that the initial frequency of the interfering gene, B, which causes the maximum reduction in $u(p_o)$ depends only on $Ni\beta$ and not on Nc. With $Ni\beta = 4$ (Figure 9) the reductions in $u(p_o)$ relative to the sampling error are much smaller, but it appears that the same conclusion holds.

Also in Figure 10 is shown a check on the theory from the diffusion equation (section 3) that the parameters $p_o$, $q_o$, $Ni\alpha$, $Ni\beta$ and Nc are sufficient to describe the system without a knowledge of the population size, N. Results for $p_o = 0.3$, $Ni\alpha = 4$ and $Ni\beta = 8$ were computed with both $N = 8$ and $N = 16$, each with 400 replicates. The data in Figure 10 has not been

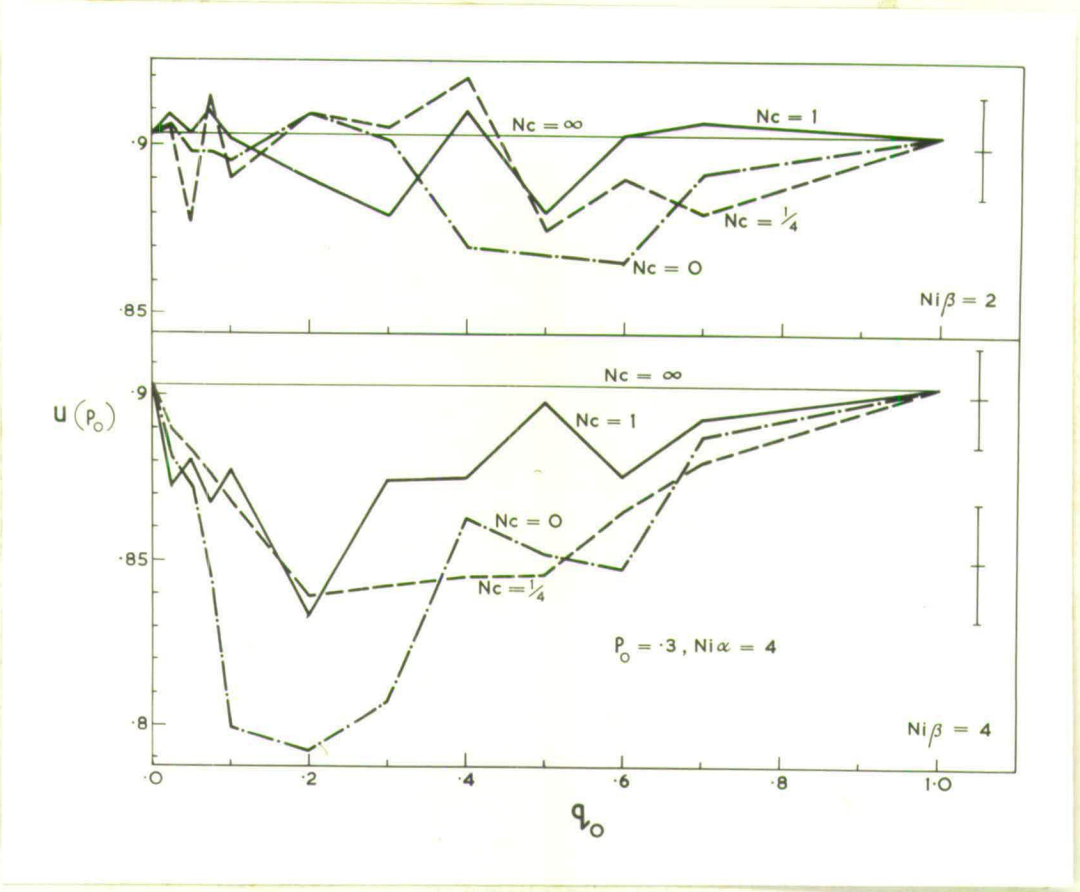


FIGURE 9 The relationship between the chance of fixation of an additive gene and the initial frequency and tightness of linkage of another additive gene with $Ni\beta = 2$ or 4. Typical ranges of length two standard deviations are also shown.

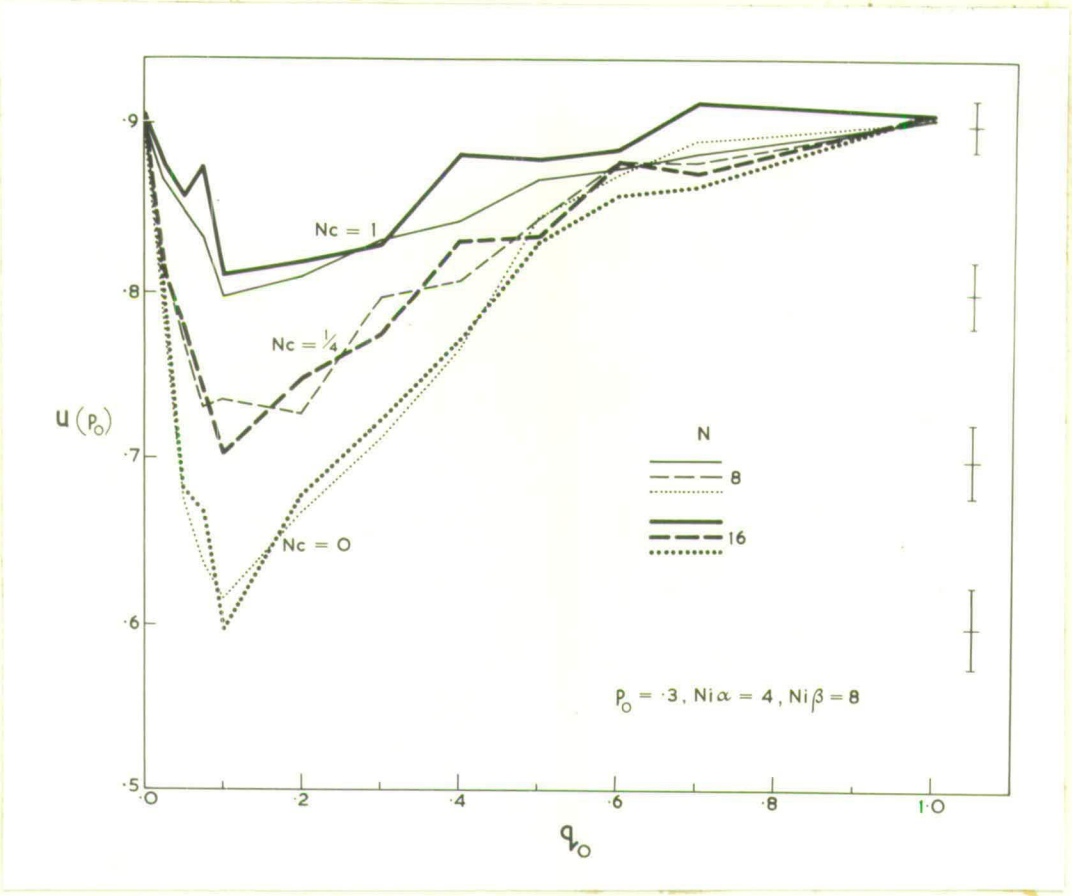


FIGURE 10 The relationship between the chance of fixation of an additive gene and the initial frequency and tightness of linkage of another additive gene with $Ni\beta = 8$. Estimates were made at two levels of population size. Typical ranges of length two standard deviations are also shown.

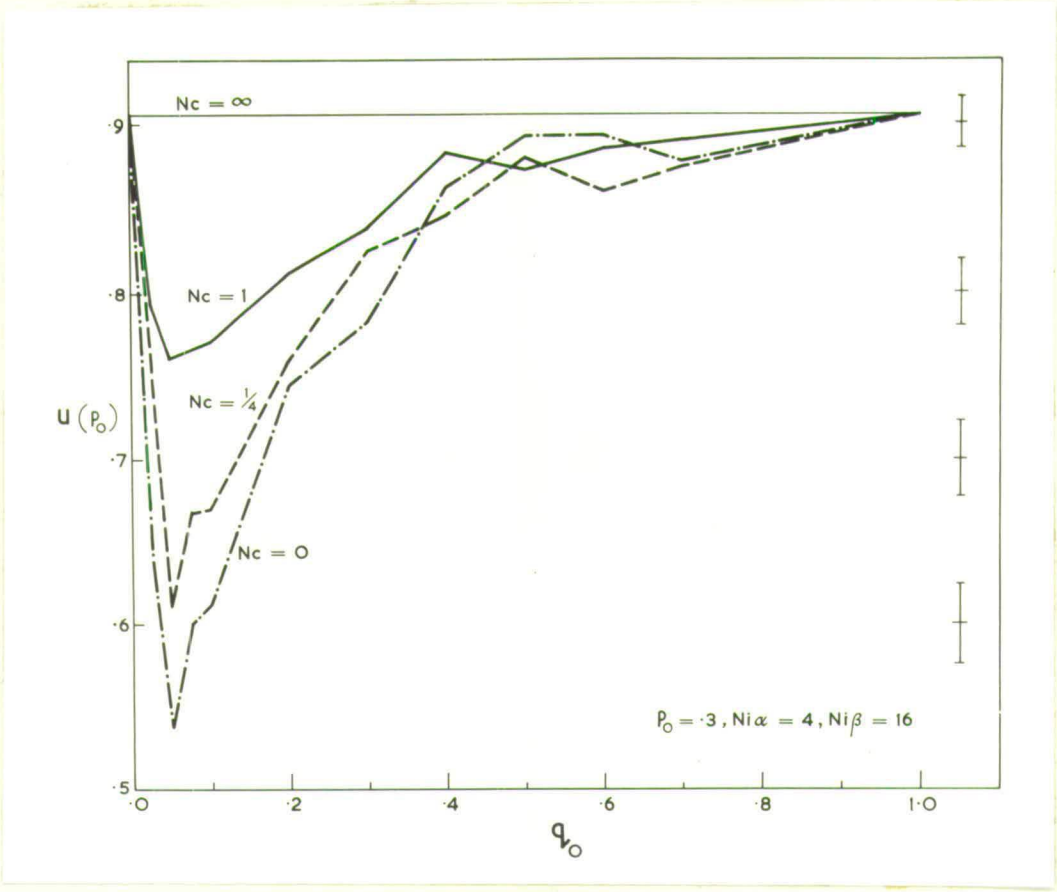


FIGURE 11 The relationship between the chance of fixation of an additive gene and the initial frequency and tightness of linkage of another additive gene with Niβ = 16. Typical ranges of length two standard deviations are also shown.

adjusted to constant population size by formula (24); the appropriate values for $u(p_o)$ are 0.9029 and 0.9056 for N = 8 and 16 respectively, when simulation is taken for 6.25 N generations.

For each value of $q_o$, the chance of fixation was compared for the run with N = 8 and N = 16. Taking each value of Nc separately, the total $\chi^2$ with 10 degrees of freedom did not differ significantly from expectation ($P > .05$) nor did the pooled $\chi^2$ with 30 degrees of freedom. The average of the differences $[u(p_o, 8) - u(p_o, 16)]$, each weighted inversely by its standard deviation, did not differ from expectation for Nc = 0 or Nc = 1/4 ($P > .05$). With Nc = 1 the average difference was found to differ significantly from zero ($.01 < P < .025$), the greater response occurring with N = 16. However, adjustment of the data by formula (24) removed this latter significant difference ($.05 < P < .1$) but did not affect any of the other comparisons.

The agreement is seen to be quite good for the example of Figure 10. This was to be expected since the values of the recombination fraction did not exceed 2/N, and thus the diffusion approximation would be expected to hold fairly well. Large values of $i\alpha$, or $i\beta$, up to 1.0 were used, but whilst these violate the diffusion equation assumptions, it turns out (Table 1) that $u(p_o)$ is not much affected by population size for this example when $Nc = \infty$.

The influence of $Ni\beta$ was studied in further detail for examples in which it was necessary to hold constant not only $p_o$ and $Ni\alpha$, but also $q_o$. The two examples studied in most detail were: $p_o = 0.1$, $q_o = 0.1$ and $Ni\alpha = 8$ (Figure 12) and $p_o = 0.3$, $q_o = 0.3$ and $Ni\alpha = 4$ (Figure 13). In each case the population sizes used were N = 8 for

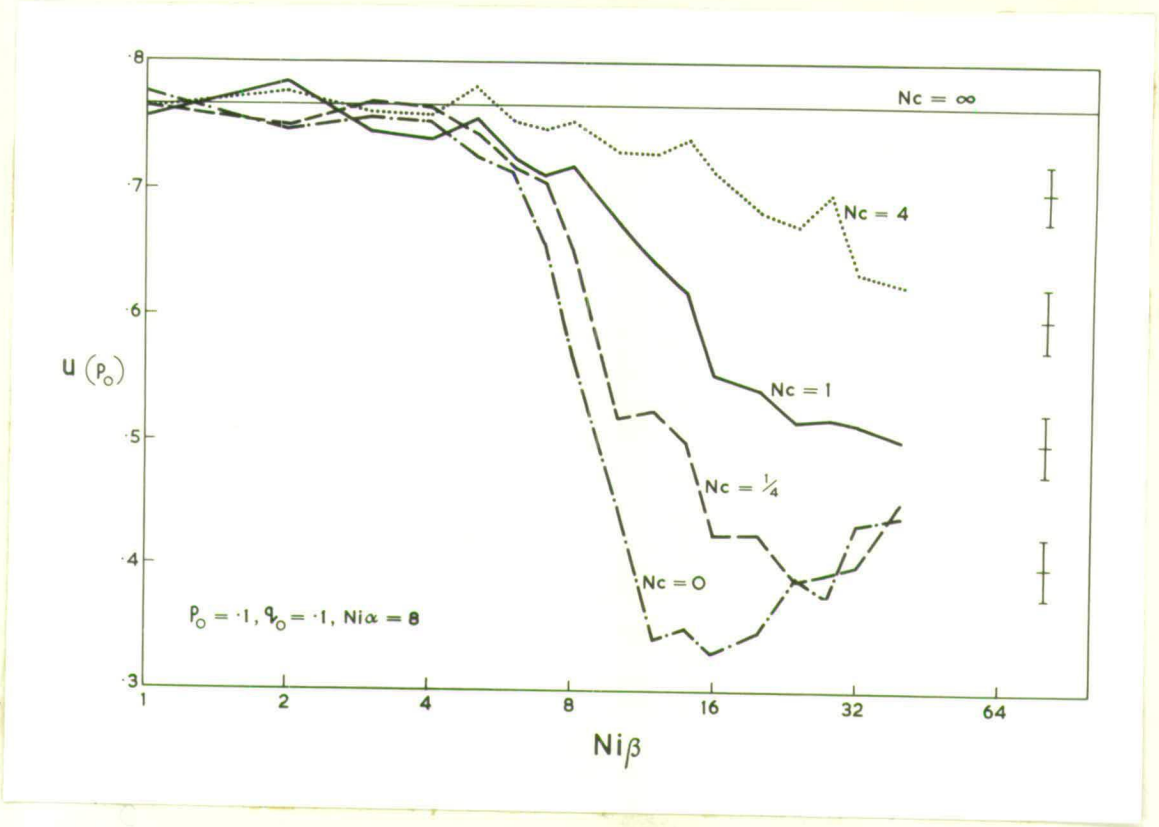


FIGURE 12 The relationship between the chance of fixation of an additive gene and the effect and tightness of linkage of another additive gene. Typical ranges of length four standard deviations if $Ni\beta \leqslant 8$ or two standard deviations if $Ni\beta > 8$ are also shown.

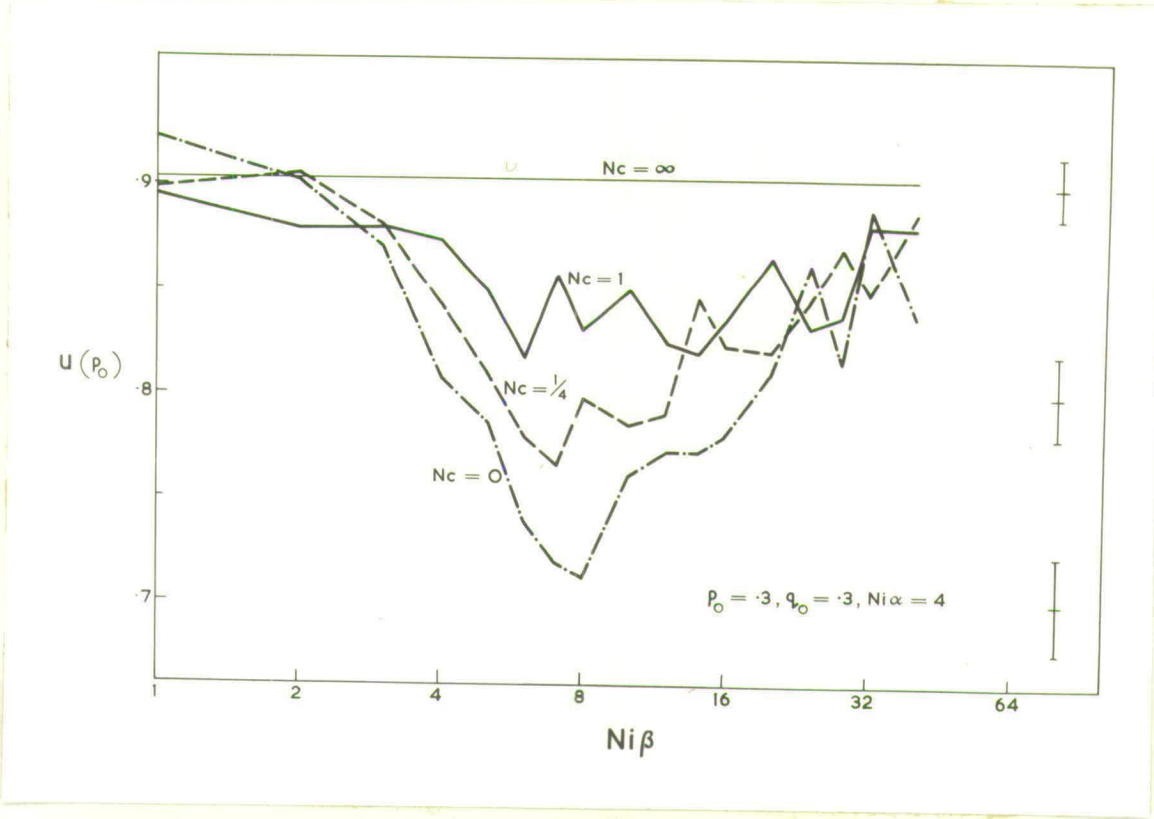


FIGURE 13 The relationship between the chance of fixation of an additive gene and the effect and tightness of linkage of another additive gene. Typical ranges of length two standard deviations are also shown.

$1 \leqslant N i\beta \leqslant 8$, $N = 16$ for $8 < N i\beta \leqslant 16$, $N = 32$ for $16 < N i\beta \leqslant 32$ and $N = N i\beta$ for $N i\beta > 32$, but all data was adjusted to $N = 8$ by formula (24). All runs were made with 400 replicates, except in Figure 12 where 1600 replicates were used for all runs in which $N i\beta \leqslant 8$. Results for $N i\beta = 0$ and $Nc = \infty$ were taken from matrix iteration.

Of particular interest in Figure 12 is the minimum value of $N i\beta$ that causes a reduction in $u(p_o)$, for in the earlier date (Figures 3-7) it was shown that little, if any, reduction occurred if $N i\beta \leqslant N i\alpha/2$, but no values of $N i\beta$ between $N i\alpha/2$ and $N i\alpha$ were run. For the parameters of Figure 12 the first significant ($P < .05$) reductions in $u(p_o)$ below that expected for an independent gene occur with $N i\beta = 5 = (5/8) N i\alpha$ when there is complete linkage, $Nc = 0$. The same relation holds, of course, between the gene effects; no reduction occurs until $\beta = (5/8)\alpha$.

A further detailed example of the fall off of $u(p_o)$ with increase in $N i\beta$ up to $N i\beta = N i\alpha$ is given in Table 3 for a model with larger effects, in which $p_o = q_o = 0.1$ as in Figure 12, but with $N i\alpha = 16$. 400 replicate runs were used.

TABLE 3 The chance of fixation of a gene with $p_o = 0.1$, $N i\alpha = 16$ and $q_o = 0.1$.

| Nc | Niβ | 0 | 2 | 4 | 8 | 10 | 12 | 14 | 16 | 32* |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | .943 | .953 | .946 | .943 | .945 | .915 | .920 | .887 | .715 |
| 1/4 | | .943 | .940 | .953 | .916 | .951 | .903 | .900 | .839 | .579 |
| 0 | | .943 | .943 | .930 | .933 | .935 | .919 | .857 | .691 | .464 |

* N = 32 adjusted to N = 16. All other runs with N = 16.

Examples of S.E. of estimates : .943 ± .012, .900 ± .015, .800 ± .020.

A significant ($P < .05$) reduction in $u(p_o)$ with $Nc = 0$ does not occur in the data of Table 3 until $Ni\beta = 12$ for which $\beta = (3/4)\alpha$. However the tests on the data in Table 3 are less powerful than those for Figure 12 since fewer replicates were used. There is also a significant reduction below that for one segregating locus in Table 3 for $Ni\beta = 8$ with $Nc = 1/4$, but a higher value of $Ni\beta$ for the same Nc does not show any change in $u(p_o)$.

For the model of Figure 13 in which $Ni\alpha = 4$ and $p_o = q_o = 0.3$ the first reduction in $u(p_o)$ is found when $Ni\beta = (3/4)Ni\alpha$ with $Nc = 0$, but fewer values of $Ni\beta$ were run. However, turning back to Figure 8 or 9, it can be seen that significant reductions occur for $Ni\beta = 2 = Ni\alpha/2$ with $Nc = 0$ where the linked gene, B, has initial frequency $q_o = 0.4$, 0.5 or 0.6. In the data of Figures 3-7 one significant reduction can be found for $Nc = 0$ and $Ni\beta = Ni\alpha/2$ for both $Ni\alpha = 8$ and $p_o = 0.1$ ($q_o = 0.5$) and $Ni\alpha = 16$ and $p_o = 0.1$ ($q_o = 0.7$), at the 5% level of significance. At the 1% level neither of these is significant and for the number of comparisons made, the 1% level of significance gives an overall Type I error of approximately 5%.

In terms of the value of $\beta$ necessary to show significant reductions in $u(p_o)$, the general conclusion that can be drawn from the data is that the critical range of values of the effect of the interfering gene is $\alpha/2 \leq \beta < 3\alpha/4$. If $\beta < \alpha/2$ no reductions have been observed, whilst if $\beta \geq 3\alpha/4$ some reduction in $u(p_o)$ seems to occur. It must be emphasised that these conclusions were drawn for interfering genes with initial frequency close to that causing the maximum reduction for its effect (Figure 8); it would be more difficult to detect deviations with

other values of $q_o$. Presumably small reductions in $u(p_o)$ occur even when $\beta < \alpha/2$, but these could not be detected with the number of replications used and are trivial relative to the order of reduction in $u(p_o)$ observed with larger $Ni\beta$ values.

When $Ni\beta$ is increased beyond $(3/4)Ni\alpha$ it can be seen in Figures 12 and 13 that $u(p_o)$ steadily declines but then passes through a minimum before increasing as $Ni\beta$ becomes much larger than $Ni\alpha$. This result could be predicted from Figure 8, where it was also shown that the value of $Ni\beta$ causing the greatest reduction in $u(p_o)$ is a function of $q_o$, the initial frequency of the interfering gene. In the example of Figure 12, it can be seen that the minimum value of $u(p_o)$ occurs at about $Ni\beta = 16$ for complete linkage, $Nc = 0$. In Table 3 the initial frequencies, $p_o = q_o = 0.1$, are the same as in Figure 12, but in the table $Ni\alpha = 16$ and in the figure $Ni\alpha = 8$. If the values of $u(p_o)$ are compared for $Nc = 0$ and $Ni\beta = 16$ and $Ni\beta = 32$ in Table 3, it can be seen that with this larger $Ni\alpha$, the maximum reduction in $u(p_o)$ occurs at an $Ni\beta$ value much greater than 16, for $Ni\beta = 32$ reduces $u(p_o)$ more than does $Ni\beta = 16$. Thus it appears that, for given $p_o$ and $q_o$, the larger $Ni\alpha$ the larger the value of $Ni\beta$ that causes the maximum reduction in $u(p_o)$. Similar results may be seen in the data of Figures 3-7. For example, with $p_o = 0.1$ (Figure 4) and $q_o = 0.3$, if the reductions in $u(p_o)$ are compared for $Ni\beta = 8$ and $Ni\beta = 16$ it is found that for the largest $Ni\alpha$ value ($Ni\alpha = 16$) a reduction in $u(p_o)$ occurs only if $Ni\beta = 16$, whereas for the smaller $Ni\alpha$ values ($Ni\alpha = 2$ or 4) a larger reduction is caused by $Ni\beta = 8$ than by $Ni\beta = 16$.

The data of Figure 12 also shows that the values of $Ni\beta$ for

which $u(p_o)$ passes through a minimum are a function of the recombination fraction. Thus for Nc = 0 and Nc = 1/4 the minima are found near Niβ = 16 and Niβ = 24, respectively, and the estimates of the chance of fixation of the gene A at these minima are $u(p_o)$ = .330 and $u(p_o)$ = .388 respectively. For Nc = 1, the maximum reduction is caused by a gene with Niβ at least 40, and from the trends of the graphs it appears that the minimum value of $u(p_o)$ lies in the range $0.45 < u(p_o) < .50$. In the example of Figure 13, the value of Niβ at the minimum of $u(p_o)$ is apparently rather less dependent on Nc than in the previous example. However the maximum reduction in $u(p_o)$ clearly depends on Nc in both examples.

In Figures 12 and 13 it can be seen that, during the phase where further increases in Niβ continue to reduce $u(p_o)$, say in the range Niβ = 6 to Niβ = 12 in Figure 12, the reduction in $u(p_o)$ is about the same whether Nc is altered from $\infty$ to 1, from 1 to 1/4 or from 1/4 to 0. In other words, during this phase, the reduction in $u(p_o)$ is linear on a scale of 1/(2Nc + 1), for given Niβ. However for larger values of Niβ, say Niβ > 2Niα in Figures 12 and 13, this scale no longer leads to linear reductions in $u(p_o)$. The reduction in $u(p_o)$ between the smaller Nc values becomes much less than between the larger values. Thus in Figure 12 there are no significant differences between $u(p_o)$ for Nc = 0 and Nc = 1/4 throughout the range Niβ = 24 to 40, the highest Niβ value simulated, whereas in this range the $u(p_o)$ values for Nc = 1 and Nc = $\infty$ differ widely. Thus the value of Niβ causing the maximum reduction in $u(p_o)$ must depend on Nc, and of course, an increase in Niβ for low Nc may increase $u(p_o)$, while reducing $u(p_o)$ for a higher Nc. An example of the

latter phenomenon can be taken from Figure 12 for the range $Ni\beta = 16$ to 40, during which $u(p_o)$ rises for $Nc = 0$ and falls for $Nc = 1$.

The detailed analysis has so far been restricted to a study of changes in the effect and frequency of the interfering gene B. The discussion will now turn to the influence of the initial frequency of the affected gene A on changes in its chance of fixation due to linkage. However, since the chance of fixation is not linearly related to the parameter $Ni\alpha$, and since the slopes of the graphs of $u(p_o)$ against $Ni\alpha$ depend on the initial frequency even for one segregating locus (Figure 1), it follows that a comparison of the changes in $u(p_o)$ itself over different values of $p_o$ does not lead to coherent conclusions. The method adopted for comparing the chance of fixation of genes with different initial frequencies was to compute from the Monte Carlo estimate of $u(p_o)$ for a linked gene the $Ni\alpha$ value which would give the same $u(p_o)$ for a single gene with the same initial frequency. This value of $Ni\alpha$, denoted $\widehat{Ni\alpha}$, was read from a graph of $u(p_o)$ against $Ni\alpha$, as in Figure 1 but using results computed by matrix iteration for the appropriate population size (Table 1). An alternative method of estimating large values of $\widehat{Ni\alpha}$ derives from a rearrangement of Kimura's (1957) formula

$$u(p_o) = \frac{1 - e^{-2Ni\alpha p_o}}{1 - e^{-2Ni\alpha}} .$$

For large $Ni\alpha$

$$u(p_o) \sim 1 - e^{-2Ni\alpha p_o}$$

so that

$$Ni\alpha \sim \frac{- \ln [1 - u(p_o)]}{2p_o} \qquad (25)$$

The approximation (25) is not satisfactory for $Ni\alpha$ 1.5, approximately, and as a large proportion of $\hat{Ni\alpha}$ values fell below 1.5 the graphical method of computing $\hat{Ni\alpha}$ was used throughout.

In Figures 14-17 $\hat{Ni\alpha}$ estimates are given for Monte Carlo runs with $Ni\beta = 16$ and $Nc = 0$ (Figure 14), $Nc = 1/4$ (Figure 15) and $Nc = 1$ (Figure 16) and with $Ni\beta = 8$ and $Nc = 0$ (Figure 17). Most of the data for these graphs was shown as chance of fixation in Figures 3-7, but there are included extra runs with 400 replicates for initial frequencies $p_o = 0.2$, 0.4, 0.6 and 0.8 with $Ni\beta = 16$, $q_o = 0.05$, 0.1 and 0.3 and $Nc = 0$. Typical sampling errors of $\hat{Ni\alpha}$ are shown in Table 4, but because $\hat{Ni\alpha}$ is not linearly related to $u(p_o)$, upper and lower bounds of $\hat{Ni\alpha}$ are shown that correspond to $u(p_o)$ plus or minus one standard deviation of $u(p_o)$.

TABLE 4 Sampling errors of $\hat{Ni\alpha}$ : upper bounds (U.B.) and lower bounds (L.B.) for $\pm$ one standard deviation of $u(p_o)$ using 400 replicates.

| $Ni\alpha$ | | $p_o$ | .05 | .1 | .3 | .5 | .7 |
|---|---|---|---|---|---|---|---|
| 1 | L.B. | | 0.77 | 0.83 | 0.89 | 0.89 | 0.87 |
| | U.B. | | 1.22 | 1.16 | 1.12 | 1.12 | 1.14 |
| 2 | L.B. | | 1.75 | 1.81 | 1.86 | 1.86 | 1.82 |
| | U.B. | | 2.25 | 2.19 | 2.15 | 2.16 | 2.27 |
| 4 | L.B. | | 3.65 | 3.73 | 3.75 | 3.68 | 3.51 |
| | U.B. | | 4.36 | 4.29 | 4.29 | 4.46 | 5.57 |
| 8 | L.B. | | 7.46 | 7.53 | 7.27 | -* | - |
| | U.B. | | 8.57 | 8.53 | 9.33 | - | - |

* Values omitted for $u(p_o) > .999$.

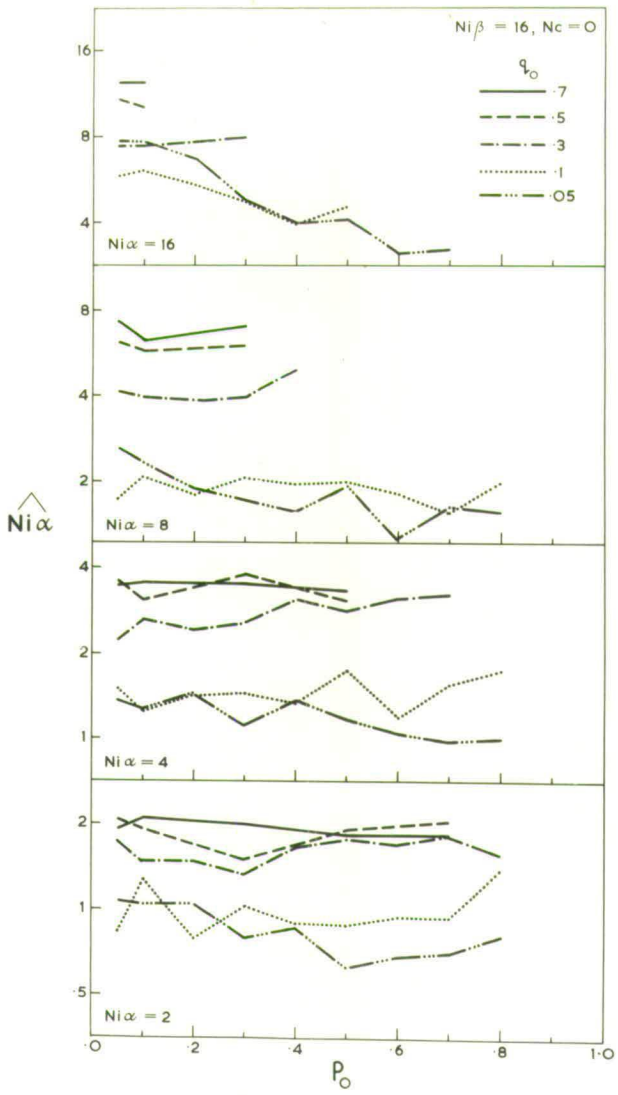


FIGURE 14 The effective selection parameter, $\widehat{Ni\alpha}$, of an additive gene as influenced by a linked additive gene with Niβ = 16 and Nc = 0.

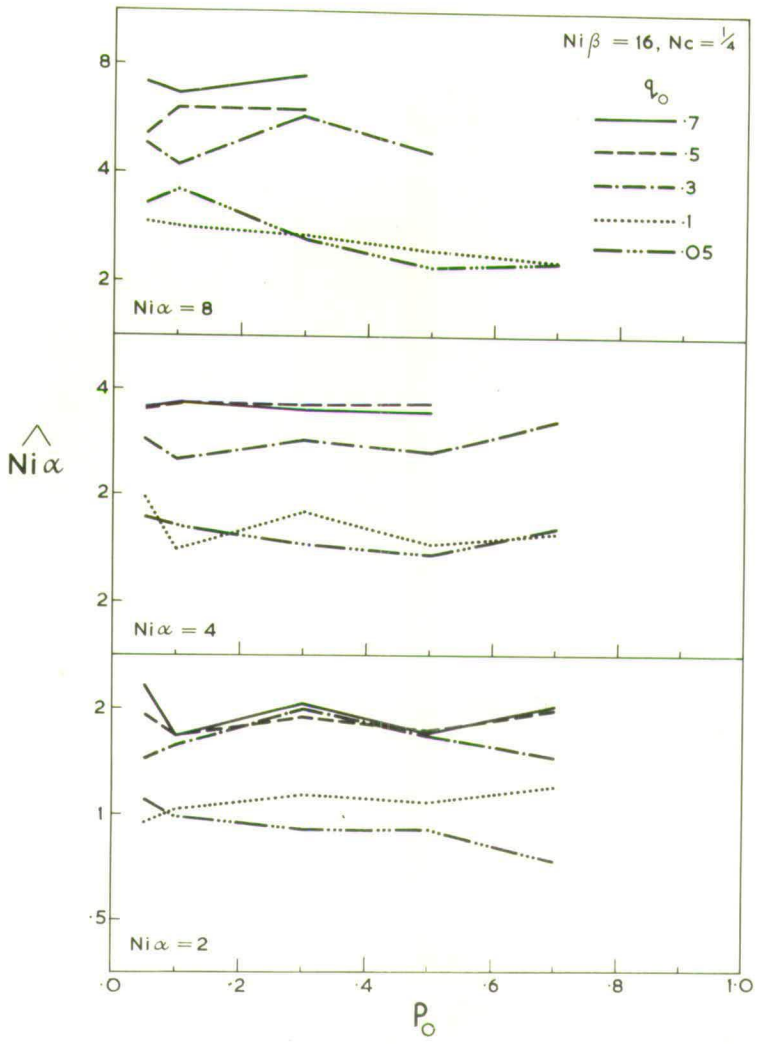


FIGURE 15 The effective selection parameter, $\hat{N}i\alpha$, of an additive gene as influenced by a linked additive gene with $Ni\beta = 16$ and $Nc = 1/4$.

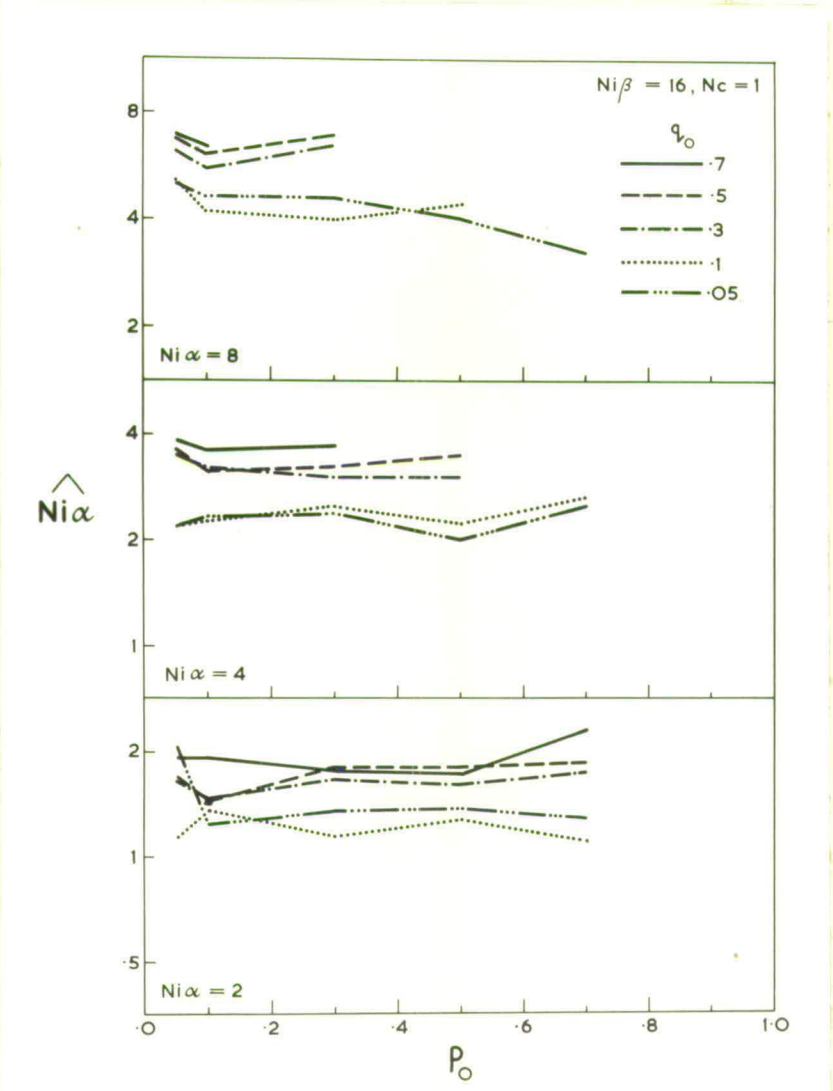


FIGURE 16 The effective selection parameter, $\hat{N}i\alpha$, of an additive gene as influenced by a linked additive gene with Niβ = 16 and Nc = 1.

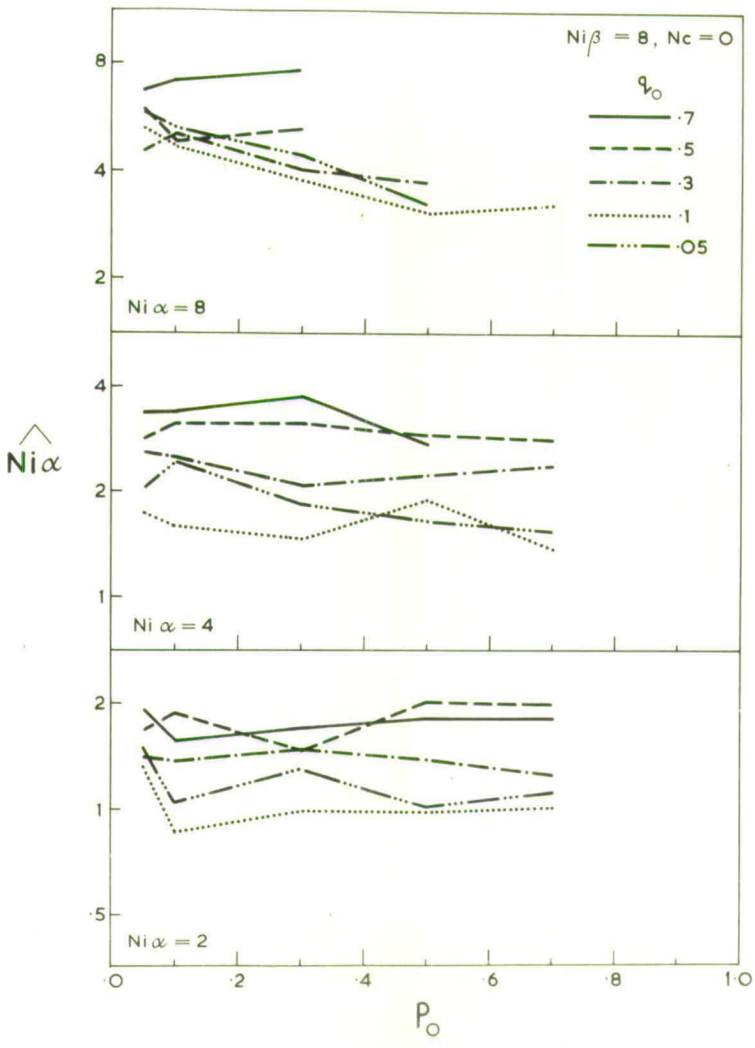


FIGURE 17 The effective selection parameter, $\widehat{Ni\alpha}$, of an additive gene as influenced by a linked additive gene with Niβ = 8 and Nc = 0.

As can be seen in Table 4, values of $u(p_o)$ close to one produce large sampling errors in $\hat{N}i\alpha$ and are also difficult to interpolate accurately from a graph, and so results for which $u(p_o) > .99$ have been omitted from Figures 14-17. Results for $Ni\alpha = Ni\beta = 16$ with $Nc = 1$ and 1/4 have not been included in Figures 15 and 16 because many $u(p_o)$ values fall above 0.99 with these parameters.

The general impression obtained from the graphs of $\hat{N}i\alpha$ is that for specific values of $Ni\alpha$, $Ni\beta$, $q_o$ and Nc the reduction in $\hat{N}i\alpha$ is approximately the same for all initial frequencies, $p_o$, of the affected gene. The pattern of reduction in $\hat{N}i\alpha$ corresponds, of course, to that for the chance of fixation : the greatest effects are caused by genes with low initial frequency, $q_o$, and with tight linkage. With high frequency genes, $q_o = 0.5$ and 0.7, $\hat{N}i\alpha$ is reduced below $Ni\alpha$ to a very small extent, and, as would be predicted from Figure 8, genes with $q_o = 0.3$ influence $\hat{N}i\alpha$ almost as much as do genes with initial frequency $q_o = 0.05$ if $Ni\beta = 8$, but to a much lesser extent if $Ni\beta = 16$.

However there are clear exceptions to the independence of $\hat{N}i\alpha$ on $p_o$. When the interfering gene has a very low initial frequency ($q_o = 0.05$), $\hat{N}i\alpha$ is reduced more if the affected gene has a high initial frequency, $p_o$. Also, if the gene effects are the same, so that $Ni\alpha = Ni\beta$, genes B of initial frequency higher than 0.05 (0.1 and perhaps 0.3) also influence $\hat{N}i\alpha$ to a greater extent if A has a high initial frequency. It has been shown that where $\alpha$ is much greater than $\beta$ (say $\alpha > 2\beta$), then the chance of fixation of the A gene is not reduced by linkage for any initial frequency, $p_o$, of A. Thus in the terms of this section, $\hat{N}i\alpha$ is not reduced for any $p_o$, given that $\alpha > 2\beta$, so that for

these relative gene effects, $\hat{N}i\alpha$ is independent of $p_o$. It is therefore probable that $\hat{N}i\alpha$ is most dependent on $p_o$ where the gene effects are the same.

Apart from the few exceptions noted above, the important conclusion that can be obtained from Figures 14-17 is that the influence of a linked gene, B, can be described solely in terms of the reduction that it causes in the selection parameter, which has been called $\hat{N}i\alpha$, without reference to the frequency of the affected gene. Thus in the earlier graphs of the influence of $q_o$, $Ni\beta$ and Nc (Figures 8-11) and of $Ni\beta$ and Nc (Figures 12 and 13), the axes showing $u(p_o)$ could be re-labelled in terms of $\hat{N}i\alpha$ and would apply to all genes of the same effect as those studied in the examples of Figures 8-13.

It was noted earlier that if $Ni\alpha$ is much greater than $Ni\beta$, no reduction in $\hat{N}i\alpha$, below the appropriate $Ni\alpha$ value, can be expected. It might be thought that for constant $Ni\beta$, then the smaller $Ni\alpha$, the greater the reduction in $\hat{N}i\alpha$, measured as a proportion of $Ni\alpha$. However, Figures 14-17 show that the reverse holds at least up to nearly equal effects; larger proportional as well as absolute changes in $\hat{N}i\alpha$, relative to $Ni\alpha$, are found with the larger $Ni\alpha$ values. For example, with $Ni\beta = 16$, $q_o = 0.1$ and $Nc = 0$ (Figure 14) estimates of $\hat{N}i\alpha$ averaged over all initial frequencies, $p_o$, are 0.99, 1.47, 1.90 and 5.1 approximately for $Ni\beta = 2$, 4, 8 and 16 respectively, and the corresponding ratios $\hat{N}i\alpha/Ni\alpha$ are 0.49, 0.37, 0.24 and 0.32 respectively. Proportional rather than absolute changes in $\hat{N}i\alpha$ clearly facilitate comparisons between wide ranges of $Ni\alpha$ values. The problem immediately raised is : what is the limiting proportion $\hat{N}i\alpha/Ni\alpha$, as $Ni\alpha$ becomes very small?

Unfortunately Monte Carlo simulation can not be expected to give very satisfactory answers, for the sampling error of $\hat{N}i\alpha/Ni\alpha$ becomes very large as $Ni\alpha$ becomes small and the curve of $u(p_o)$ against $Ni\alpha$ very flat (Figure 1). The most suitable data available has $p_o = q_o = 0.1$ and $Ni\beta = 32$, where runs with $Ni\alpha$ as low as 2, or $Ni\alpha/Ni\beta = 1/16$, have been made. For $Nc = 0$ and $Ni\alpha = 32$, 16, 8, 4 and 2 the proportions $\hat{N}i\alpha/Ni\alpha$ are 0.26, 0.20, 0.18, 0.50 and 0.60 respectively, but clearly no limit has been reached. An alternative method of finding limiting values to the proportional reduction in $\hat{N}i\alpha$ will be presented in Section 7.

Although the changes in $\hat{N}i\alpha$ have been found to be dependent on $Ni\alpha$, examination of Figures 14-17 shows that, for given $Ni\beta$, the relative influence of interfering genes of different initial frequency, $q_o$, is almost independent of $Ni\alpha$. For example, with $Ni\beta = 16$ and $Nc = 0$ (Figure 14), the reduction from $q_o = 0.5$ and $q_o = 0.7$ is nearly the same, and always small. The reduction for $q_o = 0.3$ is always less than for $q_o = 0.05$ or 0.1, except when gene effects are equal, but greater than the reduction in $\hat{N}i\alpha$ caused by genes with initial frequency $q_o = 0.5$ or 0.7. The initial frequencies $q_o = 0.05$ and 0.1 produce similar reductions in $\hat{N}i\alpha$, for all levels of $Ni\alpha$, but as mentioned previously, a gene with $q_o = 0.05$ influences genes A of higher frequency, $p_o$, rather more than genes of lower frequency $p_o$. Of course, if $Ni\alpha > 2Ni\beta$, say, no reductions for any $q_o$ would be observed. Thus, at least if $Ni\beta > Ni\alpha$, it can be concluded that the results of Figures 8-11 on the relative influence of different initial frequencies, $q_o$, hold not only for genes of different initial frequency, $p_o$, but also for genes with effects other than $Ni\alpha = 4$, the model actually studied.

It was noted earlier that when $u(p_o)$ is close to unity for Nc = the regression of $u(p_o)$ against $1/(2Nc + 1)$ is generally curvilinear. Thus the reduction in response for Nc = 0 is greater than would be expected from intermediate values of Nc if the regression were linear. Since the slope of the curve of $u(p_o)$ against $N\hat{i}\alpha$ is also strongly curvilinear when $u(p_o) \longrightarrow 1$ (Figure 1) it turns out that for these high values of $u(p_o)$, the regression of $N\hat{i}\alpha$ against $1/(2Nc + 1)$ is more closely linear than is the regression of $u(p_o)$ against $1/(2Nc + 1)$.

In this section of the thesis no attempt has been made to interpret the results from the Monte Carlo simulation on the chance of fixation of a linked gene. Some attempts at explanation of the data will be presented in Section 7, but only after further aspects of the simulation results have been discussed. Two topics are concerned with the joint chance of fixation of the two linked genes : firstly, the chance of fixation of the individual gametic types AB, Ab, aB and ab, and, secondly, the change in the population mean of some trait, where the chance of fixation of the genes or gametes must be weighted by their effects on that trait. The next section will deal with the influence of linkage on the rate of selection response during the intermediate generations before the limit is reached.

The chance of fixation of the gametes

It has been shown that the chance of fixation of each of a pair of linked genes depends upon their relative effects, initial frequencies and the tightness of linkage between them. A brief discussion will now be given on the effects of linkage on the probability of fixation

at the selection limit of each of the gametic types (AB, Ab, aB and ab with two genes each of two alleles). The case of independent segregation ($Nc = \infty$) presents no problems, for if additive genes are initially in equilibrium, then the chance of fixation of each gametic type will be the product of the chance of fixation of the genes comprising each gamete.

As an example, some results from the 1600 replicate runs with $p_o = q_o = 0.1$ and $Ni\alpha = 8$ which were used for Figure 12, are given in Figure 18 for $Nc = 4, 1, 1/4, 1/16$ and 0 and taking four examples of $Ni\beta$ : two ($Ni\beta = 2$ and 4) in which $Ni\beta$ is not more than one-half $Ni\alpha$, one ($Ni\beta = 7$) in which $Ni\beta$ is almost as large as $Ni\alpha$, and finally the case of equal effects ($Ni\beta = 8$). The chance of fixation of the favourable genes and the four gametic types are plotted against Nc, transformed to a scale of $1/(2Nc + 1)$. Previously a strict notation was adopted, in which A was termed the affected gene and B the interfering gene. Here the pair of loci are considered jointly, so the choice of label, A or B, can be made arbitrarily.

The example of Figure 18 shows that for low values of $\beta$ relative to $\alpha$ then, as mentioned previously, only the chance of fixation of the smaller gene B is reduced as linkage becomes tighter. As the gene effects become equal the fixation of both the favourable genes is reduced. On the other hand, the chance of fixation of the unfavourable coupling gamete, ab, is not influenced by the recombination fraction for any of the pair of values of $Ni\alpha$ and $Ni\beta$ shown. The chance of fixation of the repulsion gamete aB is increased only as the effect $\beta$ approaches the magnitude of $\alpha$, otherwise it is unchanged with tight linkage. The

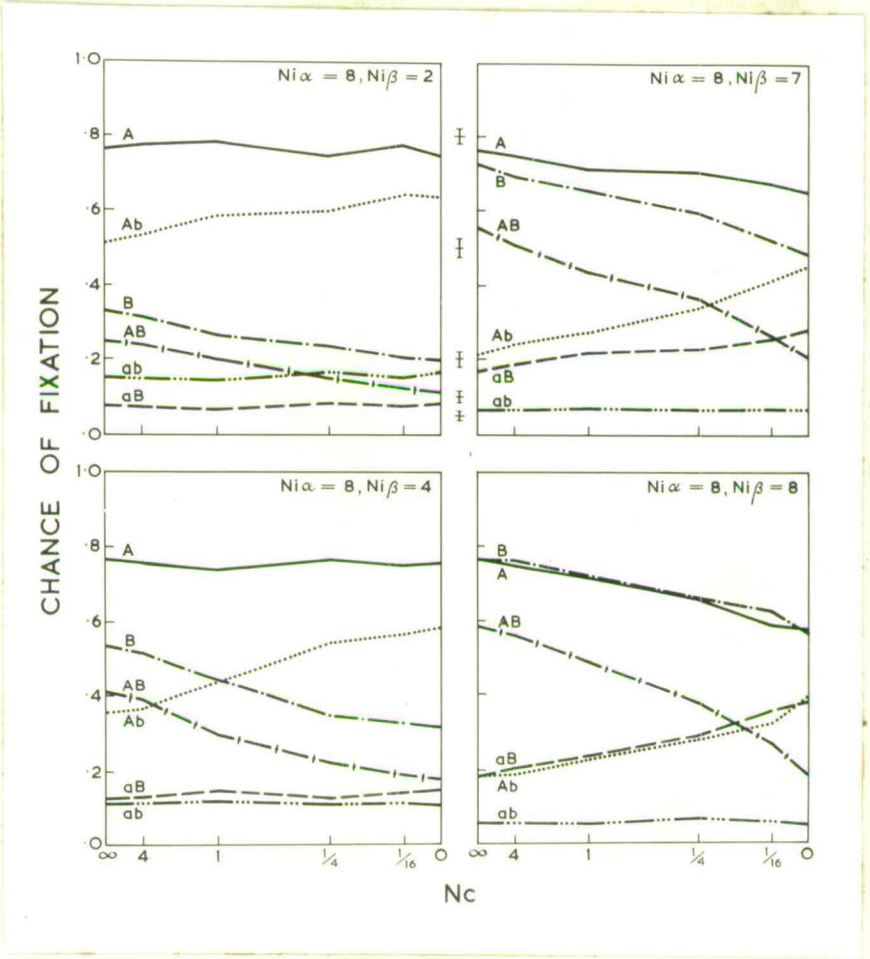


FIGURE 18 The chance of fixation of the favourable alleles and the four gametic types for a pair of linked additive genes. Typical ranges of length four standard deviations are also shown. $p_o = q_o = 0.1$.

gametes AB and Ab which contain the gene with larger effect in these examples (except where effects are equal) are influenced at all levels of $\beta \leqslant \alpha$. With tight linkage the favourable coupling gamete AB is less frequently fixed and the repulsion gamete Ab more frequently fixed, such that the sum of their frequencies, the chance of fixation of the gene A, is not affected if $\beta$ is much less than $\alpha$.

Deviations from matrix iteration results for independent loci were tested by $\chi^2$ goodness-of-fit on all the 1600 replicate runs with $p_o = q_o = 0.1$, $Ni\alpha = 8$ and $Ni\beta = 1, 2, 3, \ldots, 8$. These results confirm the impressions gained from Figure 18, for linkage was found to influence the chance of fixation in the following cases ($P < .05$) :

| | |
|---|---|
| Gene B, Gametes AB, Ab | - all $Ni\beta$ |
| Gene A | - $Ni\beta \geqslant 5$ |
| Gamete aB | - $Ni\beta \geqslant 3$ |
| Gamete ab | - no $Ni\beta$ |

Perhaps the most interesting observation that can be made from these results is that if one gene (B) is much smaller than the other, say $\beta \leqslant \alpha/4$, then the reduction in the chance of fixation of the smaller gene as linkage becomes tighter takes place only among gametes in which the large favourable gene (A) is fixed.

It can be seen in Figure 18 that as the recombination fraction becomes smaller more repulsion heterozygotes Ab, aB are fixed at the expense of coupling heterozygotes. Thus a negative linkage disequilibrium between lines at the limit, $\Delta_L$, is found, where

$$\Delta_L = u(AB) \cdot u(ab) - u(Ab) \cdot u(aB) \qquad (26)$$

and u( - ) is the chance of fixation of the specified gamete. For the examples of Figure 18 with Nc = 0, the values of $\Delta_L$ are -0.0351, -0.0667, -0.1129 and -0.1383 for Ni$\beta$ = 2, 4, 7 and 8 respectively. An excess of repulsion heterozygotes at the limit holds more generally than for the examples of Figure 18. For the runs with 400 replicates with the range of starting frequencies and effects shown in Figures 3-7, $\Delta_L$ was calculated in each of the 210 runs with Nc = 0. $\Delta_L$ was zero in 72 cases, in all of which at least one gene was fixed in all replicates, $\Delta_L$ was negative in 130 runs and positive in only 8. Moreover for none of these 8 parameter sets in which $\Delta_L$ was positive did the disequilibrium differ significantly at the 5% level from zero.

Further analyses were performed on the same data with Nc = 0 to test the other conclusions from the examples of Figure 18. Firstly, the chance of fixation of the unfavourable coupling gamete (ab) was compared with its expectation from the independent case of Nc = $\infty$ . From the 210 runs, 138 were excluded because the chance of fixation for Nc = $\infty$ fell below 0.01, so that less than 4 out of the 400 computer runs would be expected to be fixed in this class and the $\chi^2$ test can not be used. In the remaining 72 parameter sets only 2 showed significant differences ($P < .05$) between the chance of fixation of ab for Nc = 0 and Nc = $\infty$ . Secondly, for the case of unequal effects, in which $\alpha \geqslant 2\beta$ or $\beta \geqslant 2\alpha$ in this data, a similar analysis was carried out for the chance of fixation of the repulsion gamete containing the unfavourable allele of the locus with larger effect and the favourable allele of the locus with smaller effect. From the 150 comparisons of Nc = 0 with Nc = $\infty$ , 79 were excluded because the expected chance of fixation for Nc = $\infty$ was less than 0.01, and of the remaining 71

comparisons only 8 showed significant differences at the 5% probability level.

The observation that the chance of fixation of the inferior coupling gamete is not influenced by the degree of linkage therefore appears to hold generally. Since there are only three degrees of freedom among the four gametic frequencies, the additional observation that the chance of fixation of only the smaller effect gene is reduced with linkage if this gene has an effect less than, say, half that of the larger effect gene, enables prediction of the behaviour of the probability of fixation of the other three gametes as linkage becomes tighter. Thus with a wide divergence of effects, the repulsion gamete containing the favourable allele of the larger effect gene can not be influenced by linkage, as the results show. Similarly, a negative disequilibrium, $\Delta_L$, between the lines at fixation is inevitable if at least one gene has a reduced chance of fixation and the unfavourable coupling gamete's chance of fixation is not changed by tighter linkage.

The change in the population mean

The Monte Carlo results discussed so far have mostly been in terms of the chance of fixation of one gene and how this chance of fixation is influenced by another, linked, gene. In a selection experiment all that can usually be observed is the change in the population mean, which is a function of the effects and changes in frequency of all genes contributing to the trait. For two additive genes, the total advance in the mean, denoted $R(\mu)$, is given by

$$R(\mu) = \alpha\sigma [u(p_o) - p_o] + \beta\sigma [u(q_o) - q_o] \qquad (27)$$

where, as previously, $\sigma$ is the phenotypic standard deviation and the effects $\alpha$ and $\beta$ are defined as proportions of $\sigma$.

$R(\mu)$ can be calculated for each computer run from the data of Figures 3-7, but it is difficult to compare results when the initial frequencies and effects differ. Therefore the method used for making comparisons of changes in the population mean was to express the $R(\mu)$ observed for some parameter set $p_o$, $q_o$, $Ni\alpha$, $Ni\beta$, $Nc \neq \infty$ as a proportion of the response, $R(\mu)$, expected from the same parameter set, but with $Nc = \infty$. The latter results can, of course, be computed from Table 1 for it is assumed that the individual genes respond independently when $Nc = \infty$. From the data given in Figures 3-7, this proportion of the selection advance realised for $Nc = 0$ is listed in Table 5.

The greatest proportional reductions in $R(\mu)$ caused by tight linkage are found when $Ni\alpha$ and $Ni\beta$, and hence the gene effects $\alpha$ and $\beta$, are approximately equal. Such a result could be anticipated from the earlier data, for in a model in which one gene has an effect much larger than the other it has been shown that the larger gene's response is scarcely affected by the smaller linked gene. Since the gene with the larger effect generally contributes the greater part of $R(\mu)$, it therefore follows that $R(\mu)$ will also not be greatly influenced by close linkage when the genes have unequal effects.

The data from the computer runs shown in Figure 12 can also be used to illustrate that the maximum reduction in the response of the population mean caused by linkage occurs when the gene effects are approximately equal. In Figure 19 the response, $R(\mu)$, realised for

TABLE 5 Total advance of the population mean with two completely linked additive genes (Nc = 0) as a proportion of the advance expected from the same genes segregating independently (Nc = $\infty$).

| | Ni$\alpha$ | Ni$\beta$ | $q_o$ | $p_o$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | .05 | .1 | .3 | .5 | .7 |
| (a) | 16 | 16 | .05 | .659 | .671 | .777 | .885 | .914 |
| | | | .1 | .671 | .701 | .839 | .917 | .955 |
| | | | .3 | .777 | .839 | .980 | .996 | 1.000 |
| | | | .5 | .885 | .917 | .996 | 1.000 | 1.000 |
| | | | .7 | .914 | .955 | 1.000 | 1.000 | 1.000 |
| (b) | 16 | 8 | .05 | .830 | .831 | .894 | .943 | .973 |
| | | | .1 | .803 | .803 | .886 | .945 | .941 |
| | | | .3 | .806 | .869 | .958 | .986 | .992 |
| | | | .5 | .895 | .933 | .997 | .998 | 1.000 |
| | | | .7 | .920 | .939 | .998 | 1.000 | 1.000 |
| (c) | 16 | 4 | .05 | .959 | .953 | .960 | .988 | .978 |
| | | | .1 | .921 | .906 | .959 | .965 | .983 |
| | | | .3 | .932 | .923 | .963 | .995 | .984 |
| | | | .5 | .941 | .967 | .990 | .989 | .989 |
| | | | .7 | .983 | .968 | .997 | 1.001 | .999 |
| (d) | 16 | 2 | .05 | 1.006 | .983 | .996 | 1.001 | .995 |
| | | | .1 | .955 | .988 | .990 | .999 | 1.009 |
| | | | .3 | .978 | .979 | .980 | .981 | 1.000 |
| | | | .5 | .966 | .958 | .995 | .998 | .993 |
| | | | .7 | .976 | .986 | .998 | 1.001 | .995 |

TABLE 5 continued

| | Niα | Niβ | $q_o$ | $p_o$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | .05 | .1 | .3 | .5 | .7 |
| (e) | 8 | 8 | .05 | .781 | .772 | .799 | .860 | .918 |
| | | | .1 | .772 | .731 | .814 | .817 | .955 |
| | | | .3 | .799 | .814 | .880 | .947 | .988 |
| | | | .5 | .860 | .817 | .947 | .988 | .985 |
| | | | .7 | .918 | .955 | .988 | .985 | 1.000 |
| (f) | 8 | 4 | .05 | .942 | .899 | .939 | .941 | .955 |
| | | | .1 | .870 | .835 | .914 | .944 | .942 |
| | | | .3 | .863 | .887 | .892 | .964 | .989 |
| | | | .5 | .870 | .879 | .942 | .982 | .967 |
| | | | .7 | .958 | .961 | .971 | .985 | .995 |
| (g) | 8 | 2 | .05 | .930 | .963 | .990 | .985 | .995 |
| | | | .1 | 1.000 | .946 | .983 | .993 | .962 |
| | | | .3 | .993 | .881 | .971 | .961 | .971 |
| | | | .5 | .914 | .916 | .976 | 1.000 | .982 |
| | | | .7 | 1.001 | .983 | .978 | .998 | .992 |
| (h) | 4 | 4 | .05 | .775 | .868 | .846 | .883 | .906 |
| | | | .1 | .868 | .850 | .776 | .850 | .978 |
| | | | .3 | .846 | .776 | .841 | .879 | .948 |
| | | | .5 | .883 | .850 | .879 | .951 | .977 |
| | | | .7 | .906 | .978 | .948 | .977 | .982 |
| (i) | 4 | 2 | .05 | 1.053 | .965 | .982 | .992 | 1.000 |
| | | | .1 | .821 | .961 | .911 | .942 | .917 |
| | | | .3 | .881 | .859 | .907 | .905 | .958 |
| | | | .5 | .916 | .951 | .901 | .952 | .967 |
| | | | .7 | .986 | .912 | .919 | .956 | .986 |
| (j) | 2 | 2 | .05 | 1.054 | 1.064 | 1.024 | .889 | .932 |
| | | | .1 | 1.064 | .839 | .919 | .843 | .889 |
| | | | .3 | 1.024 | .919 | .826 | .917 | .899 |
| | | | .5 | .889 | .843 | .917 | .917 | .875 |

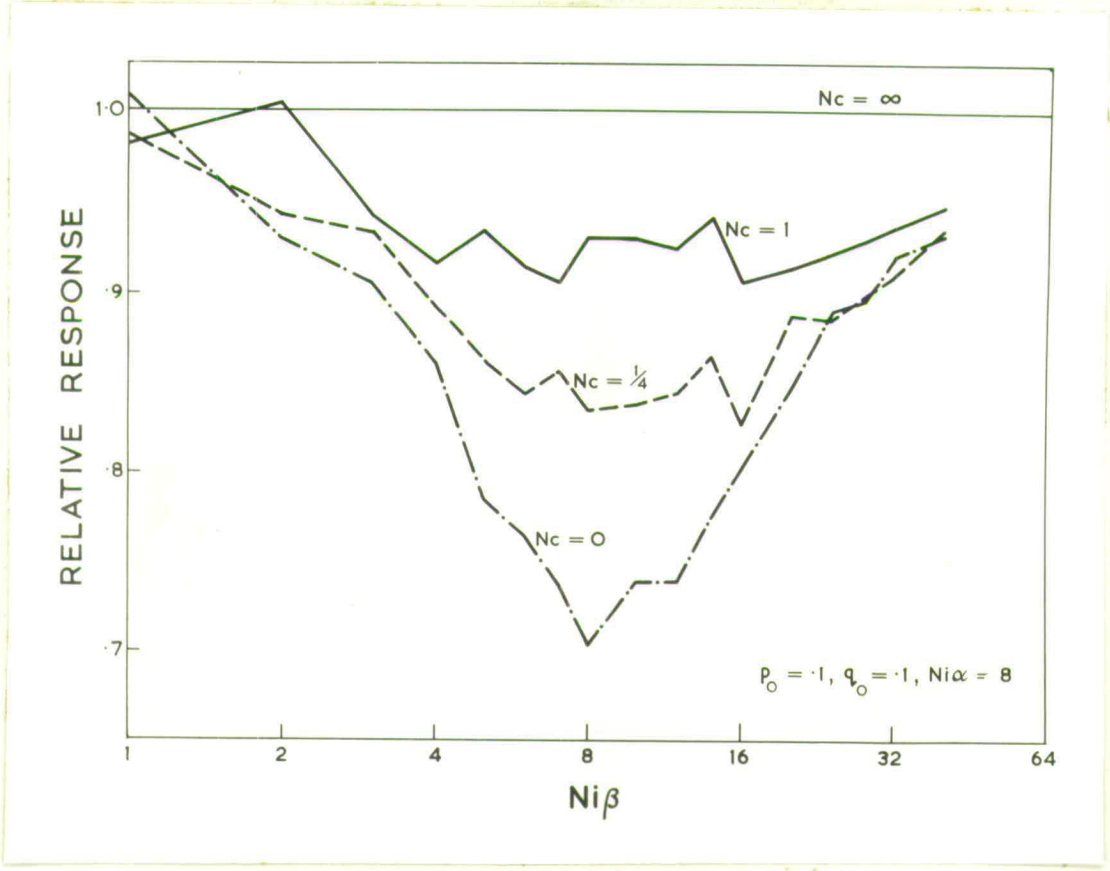


FIGURE 19 The total change in the population mean for two linked additive genes expressed as a proportion of that expected from independent genes of the same effects and initial frequencies.

Nc = 1, 1/4 and 0 is plotted as a proportion of that expected for Nc = $\infty$ with the parameters $p_o = q_o = 0.1$, Ni$\alpha$ = 8 and the wide range of Ni$\beta$ values used for Figure 12.

Turning again to Table 5, it can also be seen that the reductions in response with tight linkage occur when both genes have low initial frequency and large effects. There are two main contributing factors. Firstly, as earlier results have shown, a low frequency gene with large effect has most influence on a gene linked to it. Secondly, genes with large effect and intermediate or high frequency have a very high chance of fixation and the curve of $u(p_o)$ against Ni$\alpha$ is almost flat for such genes if Ni$\alpha$ is large (Figure 1). Therefore, even though the effective selection pressure ($\widehat{\text{Ni}\alpha}$) is reduced at least as much for high frequency genes (Figures 14-17), their response is less affected by a linked gene than is the response of a gene with low frequency and the same, large, effect.

In summary, it has been shown that the total response in the population mean for some trait determined by a pair of genes is most influenced by linkage if these genes have low initial frequency and large, approximately equal, effects.

## 6. THE RATE OF SELECTION ADVANCE

The analysis has been restricted so far to the limits of selection response. Using a few examples, a brief description will now be given of the rate of progress to the limit, with the model again restricted to a pair of additive loci initially in equilibrium. The results are relevant to all population sizes, for it has been shown (section 3) that for a given set of parameters $p_o$, $q_o$, $Ni\alpha$, $Ni\beta$ and Nc the time scale is proportional to the population size, N.

Since the approach to the limit is asymptotic, Robertson (1960) defined the half-life of the selection process, the time taken for the mean gene frequency to get half-way to the limit, as a measure of the time scale of the selection response. For one segregating locus, Robertson showed that as $Ni\alpha$ becomes very small then the half-life approximates 1.4N generations. Higher $Ni\alpha$ values usually reduce the half-life, for the favourable genes are more rapidly fixed by selection. However low frequency genes with small $Ni\alpha$ may have a half-life in excess of 1.4N generations, for the initial increase in variance due to an increase in gene frequency from selection more than compensates for the decrease in variance due to drift. Half-lives for single additive genes were computed by matrix iteration with N = 32 and are plotted in Figure 20.

Typical response curves of linked loci are shown in Figure 21. In these $p_o = q_o = 0.1$, $Ni\alpha = 2$, 4 and 8, $Ni\beta = 8$, $Nc = \infty$, 1, 1/4 and 0 and runs were made with population size N = 8. In all, 3200 replicates were used for Nc = 1, 1/4 and 0, and the curves of $Nc = \infty$ were obtained by matrix iteration. In the first few (say N/2)

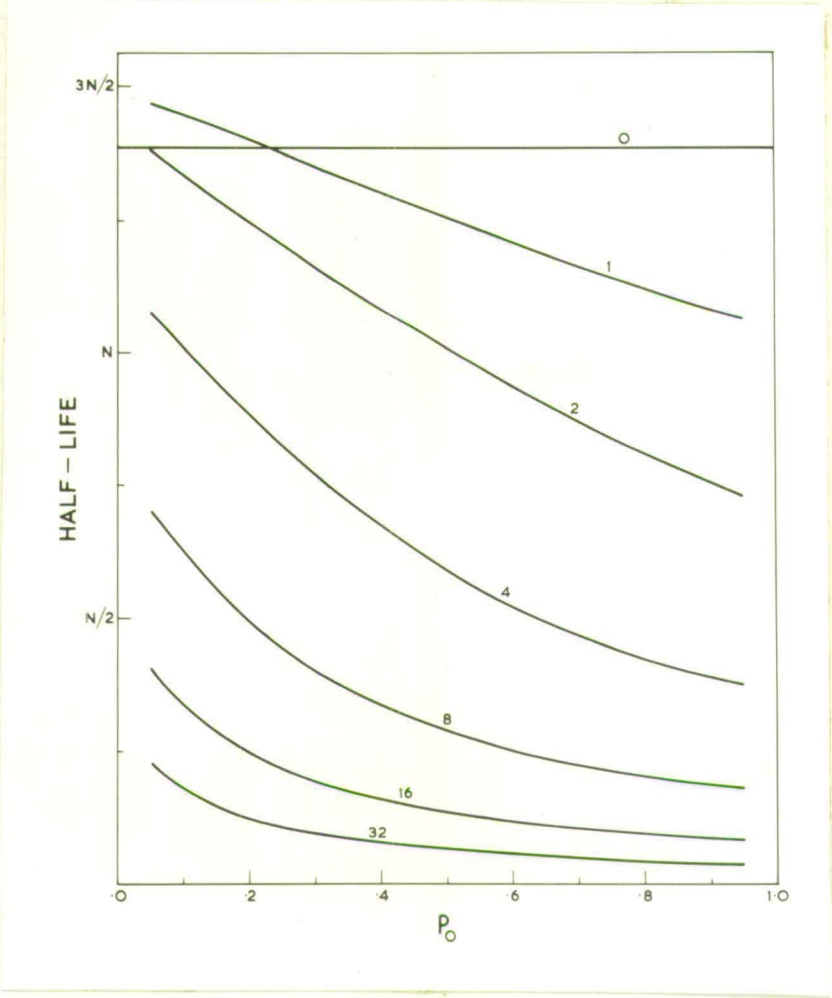


FIGURE 20 The half-life of the selection process in generations for a single additive gene. Curves are plotted for several values of Nα.

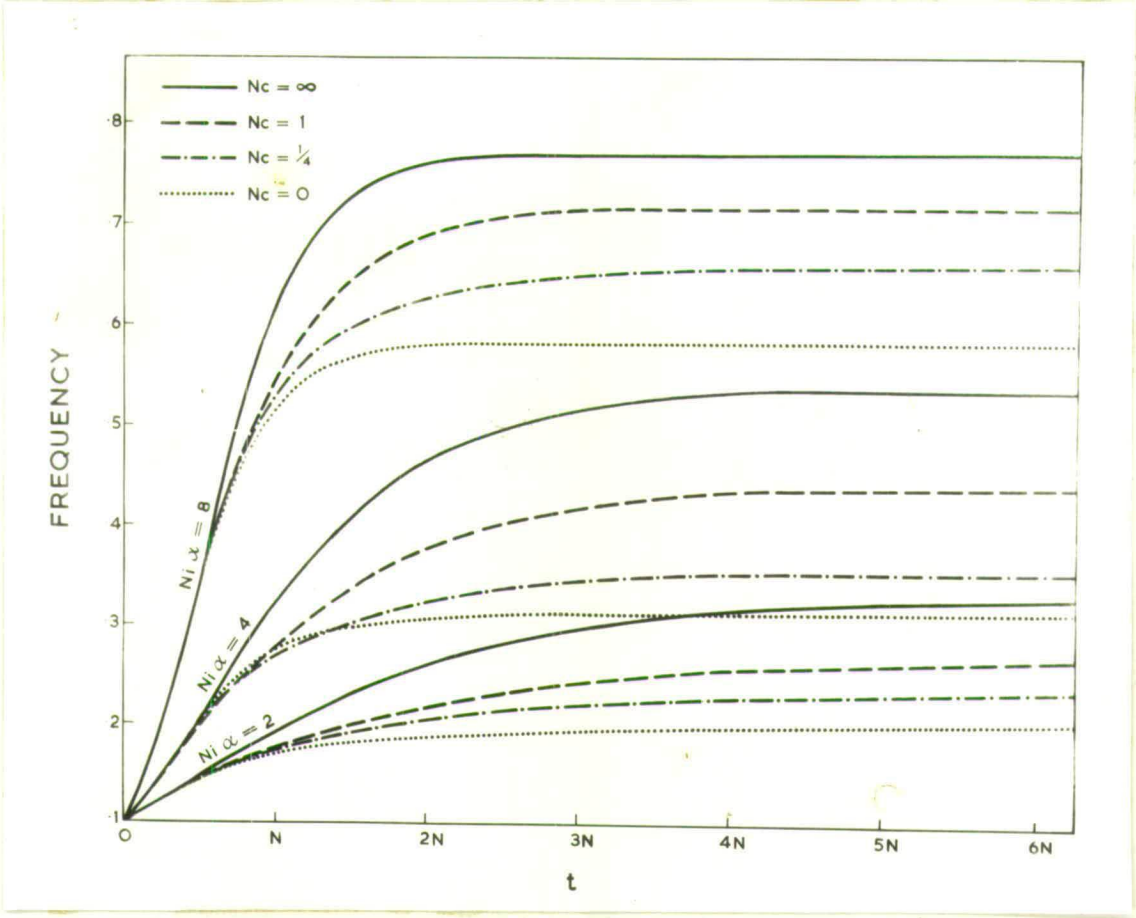


FIGURE 21 Response curves for an additive gene as influenced by its effect and tightness of linkage to another additive gene. Time is measured in generations.

generations linkage has little influence on the response but then with tight linkage the response rapidly slows down. With equal, large Niα and Niβ, it can be seen that the response ceases at about the same time with Nc = 0 and Nc = $\infty$, but some response is made later with other values of Nc as recombination takes place. Since the reduction in response with linkage only occurs in the later generations, the half-life of the selection process must be reduced. Approximate half-lives for the example of Figure 21 are given in Table 6.

TABLE 6 Half-lives (xN generations) of the selection process for $p_o = q_o = 0.1$ and Niβ = 8.

| Niα | Nc $\infty$ | 1 | 1/4 | 0 |
|---|---|---|---|---|
| 2 | 1.31 | 1.19 | 0.86 | 0.65 |
| 4 | 1.00 | 0.95 | 0.66 | 0.57 |
| 8 | 0.64 | 0.62 | 0.57 | 0.50 |

The influence of the initial frequency and size of the interfering gene, B, on the rate of response of the affected gene, A, is illustrated in Figure 22 for the example $p_o = 0.3$, Niα = 4 and Nc = 0, in which runs were made with N = 16. The selection limits (at 6.25N generations) have been shown earlier in Figures 8 and 10; they are given again in Figure 22 together with the average gene frequency of A at several intermediate stages of the selection process. The time at which response is first influenced by the linked gene, B, does not seem to depend on the initial frequency of B, but does depend on its size. If B has an initial frequency near to that causing the maximum reduction

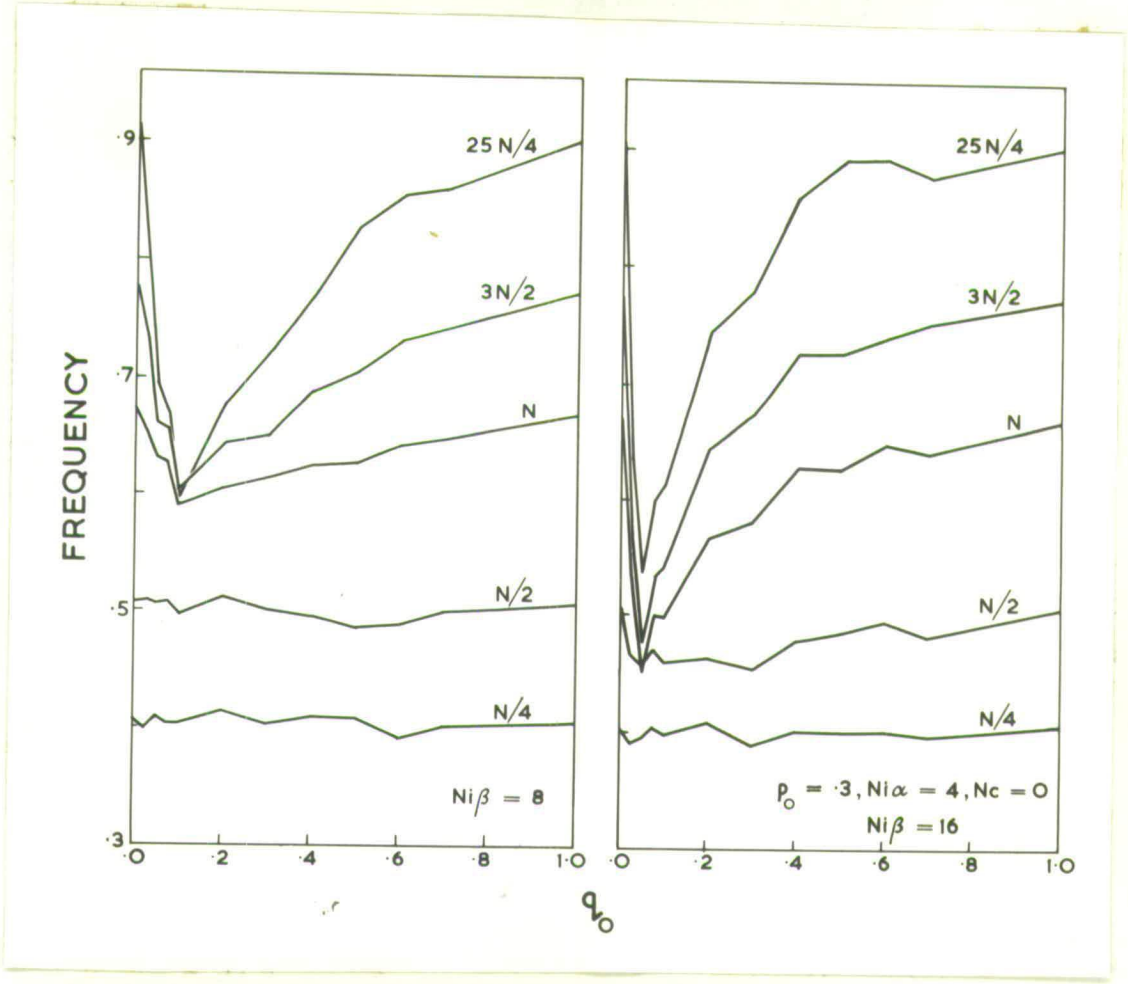


FIGURE 22 The average frequency of an additive gene after successive periods of selection, measured in generations, as influenced by the initial frequency and effect of a completely linked additive gene.

in A's chance of fixation, little further advance in the frequency of A occurs after the first effects of linkage are noted and, in fact, a small negative response may be observed for some time. If B has an intermediate or high initial frequency, it was seen previously (Figure 8) that the chance of fixation of A may be higher, the larger $Ni\beta$. Comparisons of the responses for $Ni\beta = 8$ and $Ni\beta = 16$ in Figure 22 show that it is only in the later generations that A makes a greater response with the larger value of $Ni\beta$. Approximate half-lives for this example are given in Table 7, where it can be seen that the later increase in response with the larger $Ni\beta$ and intermediate or high $q_o$ ($\geq 0.2$) is such that the half-life is actually increased somewhat by the presence of a large linked gene.

TABLE 7 Half-lives (xN generations) of the selection process for $p_o = 0.3$, $Ni\alpha = 4$ and $Nc = 0$.

| $Ni\beta$ \ $q_o$ | .0 | .025 | .05 | .1 | .2 | .3 | .4 | .5 | .6 | .7 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | .77 | .51 | .31 | .48 | .79 | .84 | .84 | .87 | .84 | .84 | .77 |
| 8 | .77 | .62 | .46 | .43 | .44 | .53 | .62 | .75 | .78 | .75 | .77 |

In Figure 23 the influence of the size of the interfering gene on the progress to the limit is studied in greater detail, using the example of Figure 13 with $p_o = q_o = 0.3$, $Ni\alpha = 4$ and $Nc = \infty$ , 1 and 0. If the results for $Nc = 0$ and $Nc = \infty$ are compared, it can again be seen that the larger $Ni\beta$ the earlier the frequency of A is influenced. Furthermore, for the highest values of $Ni\beta$, almost all the reduction in response for $Nc = 0$ versus $Nc = \infty$ occurs in these early generations. Such an

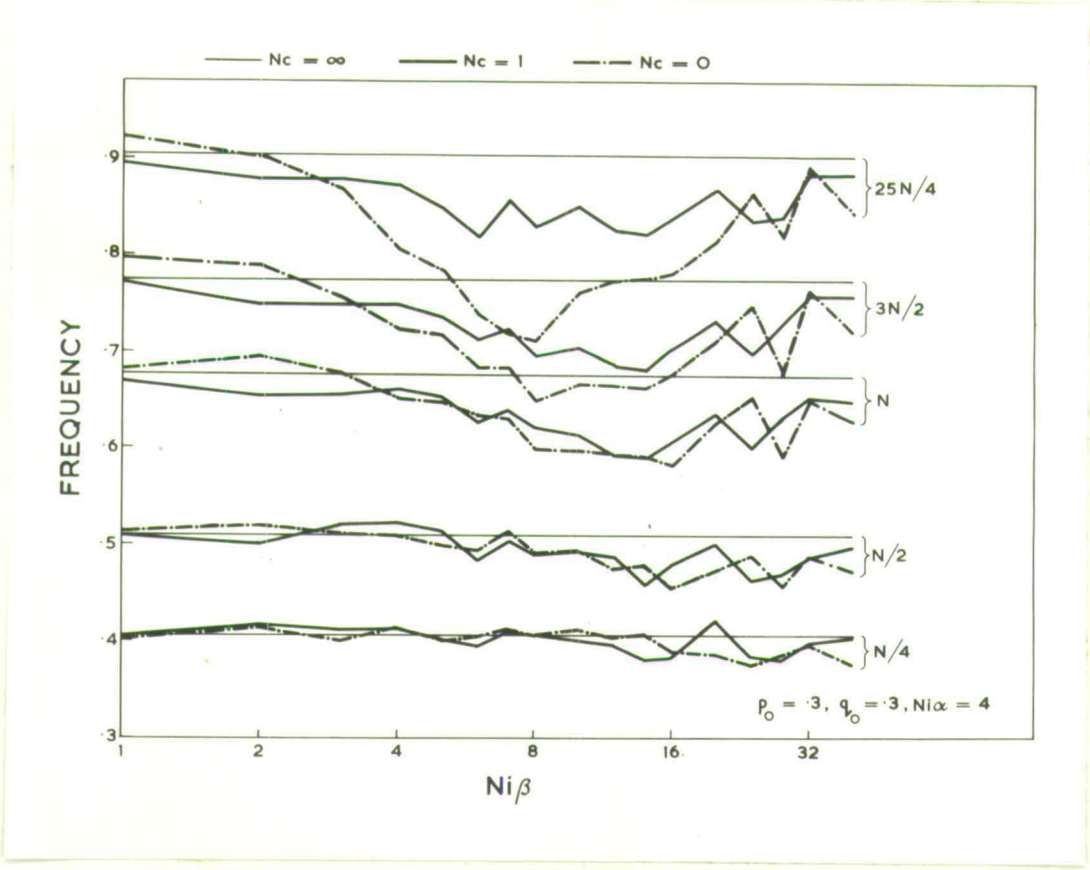


FIGURE 23 The average frequency of an additive gene after successive periods of selection, measured in generations, as influenced by the effect and tightness of linkage of another additive gene.

observation is probably to be expected, for it is only in the early generations of selection for a small gene that a very large gene remains segregating and presumably the larger gene can only affect the response while it is still segregating. For a gene with initial frequency 0.3, the .99-life, when a proportion .99 of the total response has been made, occurs after .33N, .75N, 1.79N, 3.97N and 4.61N generations for $Ni\alpha$ = 32, 16, 8, 4 and 0 respectively for a single gene, and after these times few genes remain segregating in the population.

When Nc = 1 it can be seen in Figure 22 that the initial reduction due to linkage is almost as great as when Nc = 0. It is only in the later generations and with the smaller values of $Ni\beta$, say less than 16, that more progress is made with Nc = 1 than with Nc = 0. Reasons for the build-up of linkage disequilibrium during selection will be discussed in the next section of the thesis; for the present, however, it can be assumed that differences in response for various Nc values reflect differing rates of breakdown of this disequilibrium. With the largest values of $Ni\beta$ there can be little time for recombination to occur before the large, B, gene is fixed. Thus for Nc = 1, the half-life of the breakdown of linkage disequilibrium is 0.46N generations in the absence of selection (equation 21), yet the .99-life for the selection response with $Ni\beta$ = 32 is only 0.33N generations. However, with smaller gene effects, there is more time for recombination to take place and the more closely does the rate of breakdown of linkage disequilibrium correspond to the total advance for different values of Nc.

## 7. INTERPRETATION OF SIMULATION RESULTS

In this section an attempt will be made to formulate some theory for interpreting the Monte Carlo data on selection limits and rate of response. From the theory a new method of calculating the effects of linkage on selection limits for a simple model of two loci will be developed.

### Comparison of additive and multiplicative models

It was shown earlier (10) that with the model of additive selective advantages (2), which was used in the computer programme, small amounts of negative linkage disequilibrium would be built up by selection in an infinite population initially in equilibrium. A similar effect would be expected in small populations. Negative disequilibrium has been shown to reduce the rate of selection response (7), and might therefore be expected to reduce the chance of fixation also. Such a mechanism was also proposed by Felsenstein (1965) as an explanation of the reduction in response observed by Martin and Cockerham (1960) with an additive model initially in equilibrium in small population. However, if the selective advantages $w_1$, $w_2$, $w_3$ and $w_4$ of the gametes AB, Ab, aB and ab, respectively, are multiplicative in relation to each other, such that $w_1 w_4 = w_2 w_3$, no linkage disequilibrium would result from selection in an infinite population initially in equilibrium (12). Thus, if the build-up of negative $\Delta$ shown by equation (7) is the only cause of the reduced chance of fixation with an additive model (2) in small population, no such reduction in chance of fixation would be observed with the multipli-

cative model (12). Therefore a comparison of the two models in small population was made by Monte Carlo simulation, using multiplicative selective advantages for some of the parameter sets run previously with additive selective advantages.

The multiplicative model was constructed so that if the population was in equilibrium it would remain there, and also so that changes in gene frequency would be as close as possible to those obtained with the additive model and changes due to recombination would be identical. By making these latter restrictions direct comparisons could be made between the chances of fixation obtained using the alternative models. The simulation procedure for the multiplicative model was otherwise identical with that used for the additive model (see section 4), but the changes in gametic frequencies were computed as follows:

$$\left.\begin{aligned}
AB &: f_1 + df_1 = Tf_1\,[1 + \tfrac{i}{2}\,\alpha\,(1-p)][1 + \tfrac{i}{2}\,\beta\,(1-q)] - dR \\
Ab &: f_2 + df_2 = Tf_2\,[1 + \tfrac{i}{2}\,\alpha\,(1-p)][1 - \tfrac{i}{2}\,\beta q] + dR \\
aB &: f_3 + df_3 = Tf_3\,[1 - \tfrac{i}{2}\,\alpha p][1 + \tfrac{i}{2}\,\beta\,(1-q)] + dR \\
ab &: f_4 + df_4 = Tf_4\,[1 - \tfrac{i}{2}\,\alpha p][1 - \tfrac{i}{2}\,\beta q] - dR
\end{aligned}\right\} \qquad (28)$$

where T was chosen such that $df_1 = 0$, and

$$dR = c\,\Delta\,[1 + \tfrac{i}{2}\,(\alpha + \beta - 2\mu)],$$

where $\mu = p\alpha + q\beta$, is the same as for the additive model. The change in gene frequency with (28) can be written, for A,

$$dp = \tfrac{i}{2}\,\alpha p\,(1-p) + \tfrac{i}{2}\,\beta\Delta + \frac{i^2\alpha\beta\Delta}{4+i^2\alpha\beta\Delta}\,[1 - 2p - \tfrac{i}{2}\,\alpha p(1-p) - \tfrac{i}{2}\,\beta\Delta]$$

and similarly for B. Thus the changes in gene frequency for the additive and multiplicative models differ only in terms containing both squared or cubed effects and $\Delta$ . In the initial generation both models will show the same change in gene frequency.

Computer runs with the multiplicative model were made for the parameters $p_o = q_o = 0.1$, $Ni\alpha$, $Ni\beta = 1, 2, 4, 8$ and 16 and $Nc = 4, 1, 1/4, 1/16$ and 0, all with 400 replicates. Parameter sets with $Ni\alpha = 1$ or $Ni\beta = 1$ were run with population size $N = 8$ only, those with $Ni\alpha = 16$ or $Ni\beta = 16$ were run with $N = 16$ only; all others were run with both $N = 8$ and $N = 16$. Results are shown in Figure 24 together with chances of fixation computed for the same parameters but with the additive model. Some of the latter data was also shown in Figure 4. Runs were made at different population sizes for two reasons: firstly, as a further check on the diffusion equation prediction that the chance of fixation is independent of N for each model, and secondly to test whether the comparison of the additive and multiplicative models was affected by N. Since the rate of build-up of negative disequilibrium with the additive model is a function of squared selective values (9) it seemed possible that for constant $Ni\alpha$ and $Ni\beta$ the alternative models would agree more closely at larger population sizes and correspondingly smaller selective values.



However, the general impression to be gained from Figure 24 is that the selection limit is the same for both the additive and multiplicative models at both levels of population size. Also, comparisons of the results for each model with $N = 8$ and $N = 16$ indicate that the diffusion approximation holds well. The statistical analysis

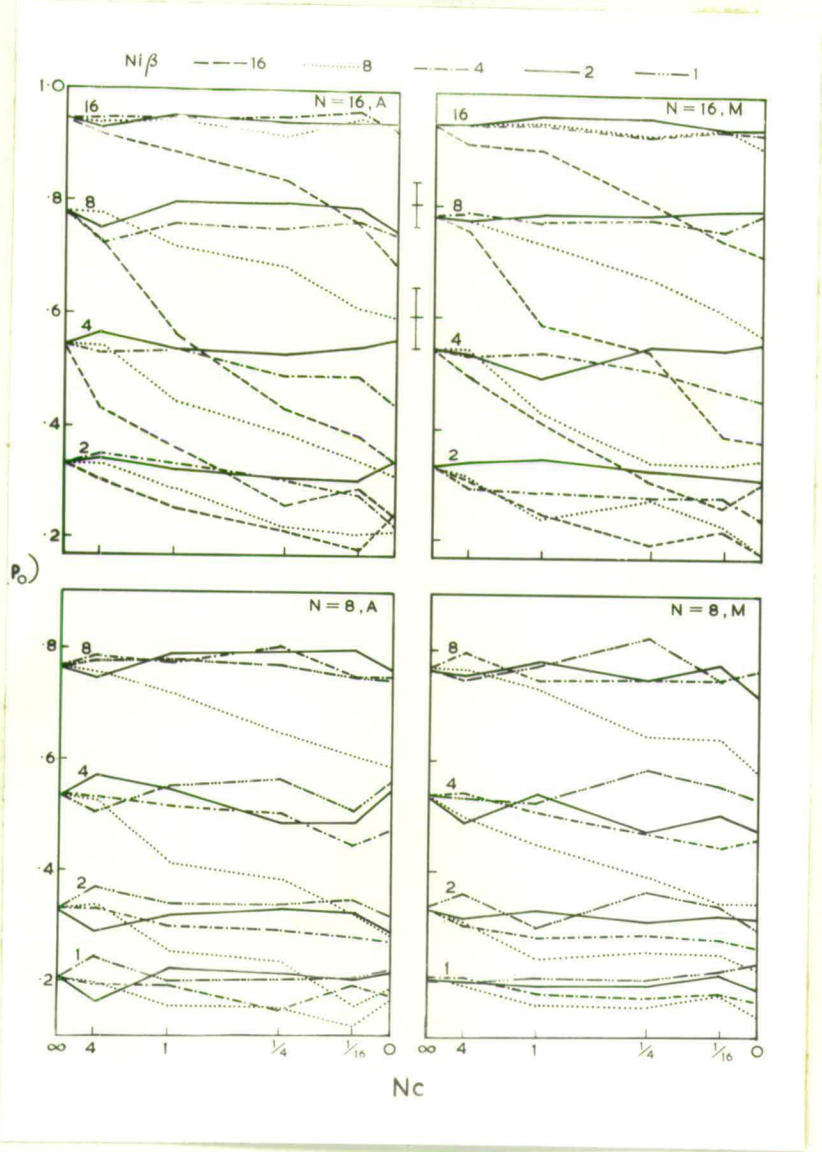


FIGURE 24. The chance of fixation of an additive gene, with various values of Niα, estimated using additive (A) and multiplicative (M) models of selective advantage each at two levels of population size. Typical ranges of length four standard deviations are also shown.

of the data of Figure 24 comprised pair-wise comparisons both of different models run with the same population size (Table 8) and of different population sizes run using the same model (Table 9). The latter analysis was made on chances of fixation both adjusted (by equation 24) to N = 8 and unadjusted. In each comparison, the difference in response for each parameter set was divided by the standard deviation of this difference and a factorial analysis performed on the new variates. In the absence of any real differences between the models each sum of squares is distributed as $\chi^2$ and has expectation equal to its degrees of freedom. Clearly the fit is good in all cases, and adjustment to N = 8 makes little improvement.

A further comparison of the additive and multiplicative models is included in Figure 25, for which the parameters were $p_o = 0.5$, $q_o = 0.1$, $Ni\alpha = 2$ and $Ni\beta = 8$. Results for $Nc = \infty$ were obtained by matrix iteration and those for $Nc = 1/2$ and $Nc = 0$ by Monte Carlo simulation with 1600 replicates and a population size of N = 8 for both models. The average gene frequency of A is plotted together with the average within-line values of $p(1 - p)$, the variance of gene frequency of A, and $\Delta$, the disequilibrium determinant or covariance of gene frequencies of A and B. From equation (7), the rate of change of gene frequency within a line, for this parameter set, can be expressed by

$$\frac{dp}{d(t/N)} = p(1 - p) + 4\Delta \qquad (29)$$

TABLE 8 Analysis of chi-square of differences in the chance of fixation for an additive vs. a multiplicative model.

| | | N = 8 | N = 16 |
|---|---|---|---|
| Source | df | Sum of squares | |
| Mean | 1 | 0.016 | 0.303 |
| Niα | 3 | 0.883 | 6.867 |
| Niβ | 3 | 10.012* | 4.725 |
| Niα x Niβ | 9 | 6.596 | 14.343 |
| Nc | 4 | 6.621 | 1.758 |
| Niα x Nc | 12 | 3.536 | 16.405 |
| Niβ x Nc | 12 | 17.566 | 7.101 |
| Niα x Niβ x Nc | 36 | 35.978 | 37.477 |
| Total | 80 | 81.208 | 88.979 |

* .01 < P < .05. All other sums of squares have P > .05.

TABLE 9 Analysis of chi-square of differences in the chance of fixation for population sizes N = 8 vs. N = 16.

| Model | | Additive | | Multiplicative | |
|---|---|---|---|---|---|
| Adjusted to N = 8 | | No | Yes | No | Yes |
| Source | df | Sum of Squares | | | |
| Mean | 1 | 3.444 | 0.003 | 3.461 | 0.000 |
| Niα | 2 | 1.929 | 6.390* | 3.439 | 1.521 |
| Niβ | 2 | 4.753 | 4.752 | 6.174* | 5.784 |
| Niα x Niβ | 4 | 2.130 | 1.186 | 3.228 | 2.824 |
| Nc | 4 | 2.691 | 2.576 | 2.268 | 2.213 |
| Niα x Nc | 8 | 5.872 | 5.916 | 7.129 | 7.426 |
| Niβ x Nc | 8 | 8.343 | 8.464 | 10.776 | 10.273 |
| Niα x Niβ x Nc | 16 | 21.153 | 20.655 | 10.492 | 10.477 |
| Total | 45 | 50.315 | 49.942 | 46.967 | 40.518 |

* .01 < P < .05. All other sums of squares have P > .05.

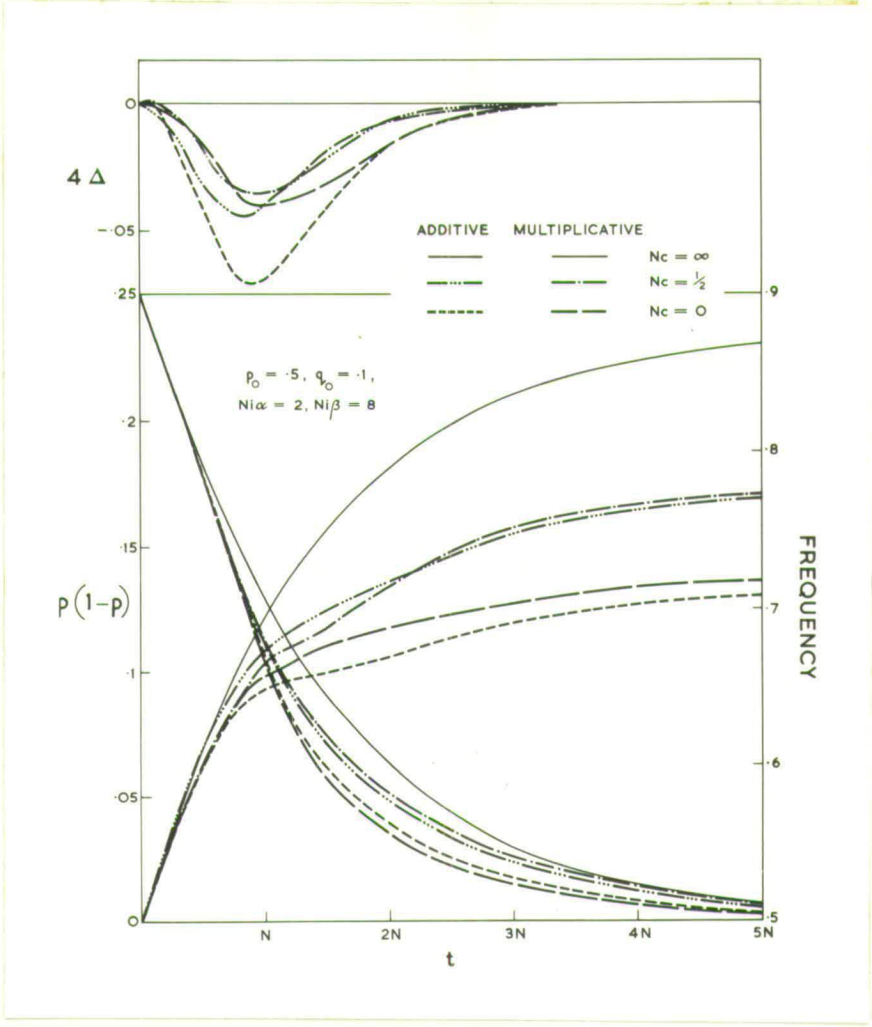


FIGURE 25 The change in the mean gene frequency, variance of gene frequency and disequilibrium determinant within lines for a linked gene using additive and multiplicative models of selective advantage. Time is measured in generations.

The scales of Figure 25 have been arranged to show the relative magnitude of the terms in (29), and curves are drawn on a time scale of t/N generations.

In Figure 25 it can be seen that with both models there is not only a build up of negative disequilibrium, but also a reduction in the variance, $p(1 - p)$, within lines when there is tight linkage between the pair of loci. The reduction in variance is about the same in both models. However, $\Delta$ becomes rather more negative with additive selective advantages, particularly in the early generations, and the chance of fixation $u(p_o)$ is slightly lower with this model, at least for the runs shown in Figure 25. Nevertheless, it is clear from both Figures 24 and 25 that the additive and multiplicative models are acting in a very similar manner, and that the build up of negative disequilibrium anticipated from (10) for the additive model in an infinite population is not an adequate explanation for the reduced chance of fixation in small populations with linked genes.

Approximately equal effects

A more satisfactory interpretation of the Monte Carlo results can be obtained by considering the sampling of gametes which occurs during selection in a small population. Firstly the discussion will be concerned with the case where the genes have approximately equal effects, but the results are rather imprecise. Then for unequal effects, say $\beta > 2\alpha$, in which the gene of smaller effect has no influence on the chance of fixation of the gene of larger effect, a more detailed approach will be presented, which leads to a useful

analytical technique.

Consider the first generation of selection of N parents from a large population in equilibrium. Let the gene frequencies in these parents be p and q for the pair of genes A and B and assume there is still equilibrium. The population of 2N sampled gametes can be classified according to which gametic types AB, Ab, aB and ab occur in the sample, as shown below.

| Class | Gametes | | | Frequency |
|---|---|---|---|---|
| | Occur | Do not occur | May occur | |
| (i) | ab | AB, Ab, aB | - | $[(1-p)(1-q)]^{2N}$ |
| (ii) | Ab,ab | AB, aB | - | $(1-q)^{2N} - [(1-p)(1-q)]^{2N}$ |
| (iii) | aB,ab | AB, Ab | - | $(1-p)^{2N} - [(1-p)(1-q)]^{2N}$ |
| (iv) | Ab,aB | AB | ab | $(1-pq)^{2N} - (1-p)^{2N} - (1-q)^{2N} + [(1-p)(1-q)]^{2N}$ |
| (v) | AB | - | Ab, aB, ab | $1 - (1-pq)^{2N}$ |

If the first sample falls into one of the classes (i), (ii) or (iii), there is at most one locus still segregating, so the degree of linkage between the pair of genes will not influence later rates of response and the chance of fixation. However in sub-populations of class (iv) both the favourable alleles A and B are represented, but there are no gametes AB. Gametes of type Ab and aB will be selected at the expense of ab, and AB will only arise as a result of recombination. If linkage is complete most sub-populations starting in class (iv) will be fixed for Ab or aB. If $\alpha = \beta$, the ratio u(Ab) : u(aB) will equal

the ratio of frequencies of Ab and aB in the initial sample. If the parameters $Ni\alpha$ and $Ni\beta$ are large, AB gametes which come from a recombination will have a high chance of fixation. With intermediate degrees of linkage, some populations initially of class (iv) may lose Ab or aB before a recombination has occurred. Thus with intermediate or tight linkage and equal large effects, a response will be seen in the early generations as the frequency of Ab and aB increase relative to ab gametes. Then a period of no response may be observed if Ab and aB have the same effect, after ab has been lost. Negative disequilibrium will, of course, be observed within lines during this period. Finally, if a recombination takes place and AB is formed, a later period of response will occur as AB increases in frequency at the expense of Ab and aB. On the average of many such populations, this phenomenon will be reflected in a long period of slow response after populations in which there is free recombination, or no recombination, have ceased to respond (Figure 21).

In sub-populations in which AB gametes are found in the initial sample (class (v)), AB will generally be fixed if $Ni\alpha$ and $Ni\beta$ are large. Otherwise, if AB is lost in the early generations, the populations will respond as for class (iv) in which AB is not found in the initial sample.

Although the above model has not been developed far enough to enable accurate predictions of chance of fixation, by making several strong assumptions some results can be obtained which indicate that the model has relevance to the observed Monte Carlo

results. Assume that there are equal effects, complete linkage and that after the first sample the selective values are sufficiently large that the most favourable gamete in each class is fixed at the limit, except in class (iv), where Ab and aB are fixed in proportion to their frequency in the sample. For an example with $p_o = 0.1$, $N = 4$ and $i\alpha = i\beta = 1.0$ for the first generation of selection, chances of fixation of the allele A were calculated for various values of $q_o$. The results were, for

$$q_o = 0.0,\ 0.05,\ 0.1,\ 0.2,\ 0.3,\ 0.5,\ 0.7,\ 0.9,\ 1.0,$$

$$u(p_o) = .714,\ .632,\ .537,\ .464,\ .479,\ .567,\ .643,\ .695,\ .714,$$

respectively. The minimum of the curve of $u(p_o)$ against $q_o$ occurs near $q_o = 0.2$, and the curve clearly resembles that of Figure 8 with $Ni\alpha = Ni\beta = 4$. The minimum of the curve occurs where the product of the probability of the initial sample falling in class (iv) and the chance that Ab is lost from this sample is a maximum.

The bottleneck model

When the interfering gene, B, has a much larger effect than A, say $\beta > 2\alpha$, a rather different approach, suggested by A. Robertson (personal communication), can be taken. Consider an additive model in which B has a low initial frequency, $q_o$, a high value of $Ni\beta$ and is completely linked to A ($Nc = 0$), such that large reductions in the chance of fixation of A are likely to be observed. Assume, firstly, that the chance of fixation of B is very close to one, a good approximation if $Ni\beta q_o > 2$. Most of the gametes fixed will therefore be AB or aB, so that all the selection response of A that is realised at

the limit will come from an increase in the frequency of AB relative to aB gametes as no recombination can take place. If, at generation t, the frequency of B is $q_t$, there will be just $2Nq_t$ gametes having the B allele so that the realised response for A will be made in a population of effective size $2Nq_t$. Thus, in the first few generations A will pass through a bottleneck of population size if B is initially at low frequency. The within line variance of A will therefore decline rapidly and less response will be made than if A is segregating independently in the population.

Consider further a model in which $Ni\alpha$ is very small, so that the mean gene frequency of A will change little as a result of selection. If there is only one additive locus, the variance declines by a proportion 1/2N each generation and the total response is given by

$$u(p_o) - p_o = \frac{i\alpha}{2} p_o (1 - p_o) \sum_{t=0}^{\infty} (1 - 1/2N)^t$$

$$= Ni\alpha p_o(1 - p_o)$$

(Robertson, 1960). For a linked additive locus, if only B gametes are fixed, then 2N must be replaced by $2Nq_t$ to describe the change in the variance and the total response, which will be denoted $\delta p_1$, is

$$\delta p_1 = \frac{i\alpha}{2} p_o (1 - p_o)[1 + (1 - \frac{1}{2Nq_1}) + (1 - \frac{1}{2Nq_1})(1 - \frac{1}{2Nq_2}) + \ldots] \quad (30).$$

However, not all the gametes fixed will, in general, contain the favourable allele B, so the theory must be developed further to include this situation. Further, it will be necessary to calculate

the distribution of gene frequencies, $q_t$, in order to evaluate (30).

Let $p_1$ be the frequency of A alleles among gametes having the B allele, i.e. $p_1 = f_1/(f_1 + f_3)$ in terms of gametic frequencies, and let $p_2$ be the frequency of A alleles among gametes having the b allele. Therefore

$$p = p_1\, q + p_2\, (1 - q). \qquad (31)$$

At the limit, letting $p_1$ and $p_2$ be the total changes in frequency of A among B and b gametes respectively, the average chance of fixation of A can be written in the same form as (31) as

$$u(p_o) = (p_o + \delta p_1)\, u(q_o)^* + (p_o + \delta p_2)[1 - u(q_o)^*], \qquad (32)$$

where $u(q_o)^*$ is the chance of fixation of B, which must be evaluated. The disequilibrium determinant can be written as

$$\Delta = (p_1 - p_2)q(1 - q)$$

so equation (7) for the change in frequency of B becomes

$$dq = \frac{i\beta}{2}\, q(1 - q) + \frac{i\alpha}{2}\, (p_1 - p_2)q(1 - q)\ . \qquad (33)$$

Kimura's (1957) formula (equation 17) for the chance of fixation holds for one locus, and is thus comprised only of the term $\frac{i\beta}{2}\, q(1 - q)$ in (33). In the model here, it is assumed that $i\alpha$ is very small, so that changes in q due to disequilibrium and selection on A will also be small. Thus if $\frac{du(q)}{dq}$ is the differential of changes in the chance of fixation of B with changes in gene frequency,

$$du(q)_A = [\frac{du(q)}{dq}]dq_A$$

where $du(q)_A$ and $dq_A = \frac{i\alpha}{2}\, (p_1 - p_2)q(1 - q)$ are the changes in chance of fixation and gene frequency of B due to selection on A. From equation (17)

$$\frac{du(q)}{dq} = \frac{2Ni\beta\, e^{-2Ni\beta q}}{1 - e^{-2Ni\beta q}} \quad . \qquad (34)$$

If $u(q_o)$ is the chance of fixation of B computed for the one locus model, and $\delta u(q_o)_A$ is the adjustment that must be made to take account of selection for A, then

$$u(q_o)^* = u(q_o) + \delta u(q_o)_A .$$

Equation (32) can then be rewritten to give the response in A as

$$u(p_o) - p_o = (\delta p_1)u(q_o) + (\delta p_2)[1 - u(q_o)] + (\delta p_1 - \delta p_2)\,\delta u(q_o)_A, \qquad (35)$$

where account has yet to be taken of the distribution of frequency of B.

If the changes in $p_1$ and $p_2$ at generation T are denoted $dp_{1T}$ and $dp_{2T}$ respectively, then

$$\delta p_1 = \sum_{T=0}^{\infty} dp_{1T} \quad , \qquad \delta p_2 = \sum_{T=0}^{\infty} dp_{2T} .$$

Also, at generation t,

$$dq_A = \frac{i\alpha}{2}\Big[\sum_{T=0}^{t} (\delta p_{1T} - \delta p_{2T})\Big]\, q_t\,(1 - q_t).$$

Thus

$$(\delta p_1 - \delta p_2)\,\delta u(q_o)_A = \frac{i\alpha}{2} \sum_{t=0}^{\infty} \Big[\sum_{T=0}^{\infty} (dp_{1T} - dp_{2T})\Big]\Big[\sum_{T=0}^{t} (dp_{1T} - dp_{2T})\Big]$$

$$\cdot\; q_t\,(1 - q_t)\,\frac{du(q)}{dq} \quad . \qquad (36).$$

The terms $dp_{1T}$ or $dp_{2T}$ in (36) each have a component from drift, which has zero expectation, and a component from selection. Products of two such terms will include a function of $(i\alpha)^2$ which can be ignored as this is of smaller order of magnitude than the other components of (35). So that

$$E(dp_{1T}dp_{2T'}) = 0 \quad \text{, all } T, T',$$

$$E(dp_{1T}dp_{1T'}) = 0 \text{ and } E(dp_{2T}dp_{2T'}) = 0 \quad \text{, } T \neq T',$$

$$E(dp_{1T}^2) = V(dp_{1T}) \text{ and } E(dp_{2T}^2) = V(dp_{2T}),$$

where V denotes variance. Hence (36) gives

$$(\delta p_1 - \delta p_2)\,\delta u(q_o)_A = \frac{i\alpha}{2} \sum_{t=0}^{\infty} \left\{ \sum_{T=0}^{t} [V(dp_{1T}) + V(dp_{2T})] q_t(1 - q_t) \frac{du(q)}{dq} \right\}. \quad (37)$$

In order to find the expectation of the terms in (35), it will be necessary to investigate the gene frequency distribution of B. To do this it is convenient to define some matrices at the outset. Let $M^*$, M, X and Y be square matrices of dimension 2N + 1, with elements $m^*_{jk}$, $m_{jk}$, $x_{jk}$, $y_{jk}$, respectively, for j,k = 0, 1, ..., 2N, where:

$$m^*_{jk} = \binom{2N}{k} \left[\frac{j}{2N} + \frac{i\beta}{2} \cdot \frac{j}{2N}\left(1 - \frac{j}{2N}\right)\right]^k \left[1 - \frac{j}{2N} - \frac{i\beta}{2} \cdot \frac{j}{2N}\left(1 - \frac{j}{2N}\right)\right]^{2N-k} ;$$

$$m_{jk} = m^*_{jk} \text{ all } j, k \text{ except } m_{oo} = m_{2N,2N} = 0 ;$$

$$x_{jk} = m^*_{jk}(1 - 1/k) \text{ , } k \neq 0 ; \; x_{jo} = 0 ;$$

$$y_{jk} = m^*_{jk}\left(1 - \frac{1}{2N-k}\right), \; k \neq 2N ; \; y_{j,2N} = 0 .$$

$M^*$ is, of course, the transition probability matrix for the change in gene frequency, where $j/2N = q_t$ , $k/2N = q_{t+1}$. Also let U, L and V be column vectors of dimension 2N + 1, with elements $u_j$, $l_j$ and $v_j$, respectively, for j = 0, 1, ..., 2N, where:

$$u_j = \frac{1 - e^{-ji\beta}}{1 - e^{-2Ni\beta}} \text{ , the chance of fixation of B ;}$$

$$l_j = 1 - \frac{1 - e^{-ji\beta}}{1 - e^{-2Ni\beta}}$$ , the chance of fixation of b;

and $v_j = i\beta j(1 - j/2N)\left(\frac{1 - e^{-ji\beta}}{1 - e^{-2Ni\beta}}\right)$, which in terms of $q_t$ is

$$v_j = q_t(1 - q_t)\frac{du(q)}{dq} .$$

Finally let $R_t$ be a column vector of the same dimensions, with elements $r_{j(t)}$. At generation zero, $R_o$ defines the initial gene frequency distribution of B, and if $2Nq_o = j$ is integral, $R_o$ becomes

$$r_{k(o)} = 0,\ k \neq j \quad \text{and} \quad r_{j(o)} = 1. \tag{38}$$

Turning firstly to selection among gametes having the favourable allele B, imagine that at generation (t-1) the frequency of B is j/2N and the variance in gene frequency of A among these gametes is $p_o(1 - p_o)\, r_{j(t-1)}$. At generation t, B will have frequency k/2N with probability $m_{jk}$, in which case the variance will decline by a proportion 1/k. The expected value of $r_{k(t)}$ is therefore

$$r_{k(t)} = \sum_{j=1}^{2N} r_{j(t-1)} m_{jk} (1 - 1/k)$$

$$= \sum_{j=1}^{2N} r_{j(t-1)}\, x_{jk}. \tag{39}$$

Summation in (31) does not have to be taken over j, k = 0 since B is not segregating. However, in order to keep all matrices with the same dimensions, it is convenient to do so. Since, by definition $x_{jo} = 0$, (39) can be replaced by

$$r_{k(t)} = \sum_{j=0}^{2N} r_{j(t-1)}\, x_{jk} \quad . \qquad (40)$$

In matrix notation (40) becomes

$$R'_t = R'_{t-1} X$$

$$= R'_o X^t$$

where $X_o = I$, the identity matrix. X can be regarded as a transition matrix for changes in the variance of A. From (35) the contribution from generation t to $u(p_o)-p_o$ due to selection within B gametes is seen to be the expectation of $dp_{1t}u(q_o)$. This is

$$E[dp_{1t}\, u(q_o)] = \frac{ia}{2} p_o(1 - p_o) \sum_{k=0}^{2N} r_{k(t)}\, u_k$$

$$= \frac{ia}{2} p_o(1 - p_o)\, R'_o X^t U.$$

Therefore, the total response is

$$E[\delta p_{1t}\, u(q_o)] = \frac{ia}{2} p_o(1 - p_o)\, R'_o \left(\sum_{t=0}^{\infty} X^t\right) U \quad . \qquad (41)$$

It can be shown that all the latent roots of X, $\lambda_j$, are such that $|\lambda_j| < 1$, so that the following relation holds

$$\sum_{t=0}^{\infty} X^t = (I - X)^{-1}$$

(e.g. Finkbeiner, 1960, p. 196). Hence from (33)

$$E[\delta p_{1t}\, u(q_o)] = \frac{ia}{2} p_o(1 - p_o)\, R'_o (I - X)^{-1} U \quad . \qquad (42)$$

Among gametes in which b is finally fixed, the variance each generation declines by a proportion $\frac{1}{2N(1-q_t)}$. The expected total response from selection among b gametes is therefore given by

$$E\left\{\delta p_2\,[1 - u(q_o)]\right\} = \frac{ia}{2}\, p_o(1 - p_o)\, R_o'\,(I - Y)^{-1}L \quad . \qquad (43)$$

Finally consider the expectation of the remaining term in (35), which is $(\delta p_1 - \delta p_2)\,\delta u(q_o)_A$ and is expanded in (37). The expectation of $\sum_{T=0}^{t} V(dp_{1T})$ will be the drift variance among B alleles, and will be

$$E[\sum_{T=0}^{t} V(dp_{1T})] = p_o(1 - p_o)[\frac{1}{2Nq_1} + \frac{1}{2Nq_2}(1 - \frac{1}{2Nq_1}) + \ldots + \frac{1}{2Nq_t}\prod_{T=1}^{t}(1 - \frac{1}{2Nq_T})], \qquad (44)$$

since changes in $p_o$ are assumed to be small. (44) can be rearranged to give

$$E[\sum_{T=0}^{t} V(dp_{1T})] = p_o(1 - p_o)[1 - \prod_{T=1}^{t}(1 - \frac{1}{2Nq_T})] \quad .$$

Therefore from (37)

$$(\delta p_1 - \delta p_2)\,\delta u(q_o)_A = \frac{ia}{2}\, p_o(1 - p_o) \sum_{t=0}^{\infty}\Big\{[2 - \prod_{T=1}^{t}(1 - \frac{1}{2Nq_T}) - \prod_{T=1}^{t}(1 - \frac{1}{2N(1-q_T)})]\, q_t(1 - q_t)\frac{du(q)}{dq}\Big\}. \qquad (45)$$

From the transition matrix for the gene frequency distribution, M, and from the derivations used to evaluate the responses within gametes containing B or b alleles, the expectation of (45) becomes

$$E[(\delta p_1 - \delta p_2)\,\delta u(q_o)] = \frac{ia}{2}\, p_o(1 - p_o)\, R_o' \sum_{t=0}^{\infty}(2M^t - X^t - Y^t)V$$

$$= \frac{ia}{2}\, p_o(1 - p_o)\, R_o' \,[2(I-M)^{-1} - (I-X)^{-1} - (I-Y)^{-1}]V. \qquad (46)$$

Summing (42), (43) and (46) gives the final result for the expected response,

$$u(p_o)-p_o = \frac{i\alpha}{2} p_o(1-p_o) R_o' [2(I-M)^{-1}V + (I-X)^{-1}(U-V) + (I-Y)^{-1}(L-V)] \quad (47)$$

Formula (47) was used to compute responses, $u(p_o)-p_o$, for a range of $Ni\beta$ values, and results are shown as a function of $q_o$, the initial frequency of the interfering gene, B, in Figure 26. The responses are plotted as a proportion of $Ni\alpha\ p_o(1 - p_o)$, which is the total response expected for a single gene of the same effect and initial frequency (Robertson, 1960), and so are independent of $Ni\alpha$ and $p_o$. In the notation used in section 5, this ratio is therefore $\hat{Ni\alpha}/Ni\alpha$.

If Figure 26 is compared with Figure 8, in which $u(p_o)$ is plotted against $q_o$ for various $Ni\beta$ and $Nc = 0$ from Monte Carlo runs, it can be seen that there is a striking resemblance between the graphs. The agreement is best for the higher $Ni\beta$ values ($Ni\beta \geqslant 8$), where in Figure 26 it is again found that the maximum reduction in the chance of fixation of A occurs when $q_o = 0.8/Ni\beta$, approximately, for which $u(q_o) = 0.8$, approximately. A larger maximum reduction in $u(p_o)$ for $Ni\beta = 32$ relative to $Ni\beta = 16$ is found in Figure 26 but not in Figure 8. It seems probable that the contrary observation from the Monte Carlo run was caused by sampling. However, since the gene A studied by Monte Carlo simulation for Figure 8 has $Ni\alpha = 4$, little, if any, reduction in $u(p_o)$ is found where $Ni\beta \leqslant 2$ or, in terms of the effects, $\beta \leqslant \alpha/2$. The bottleneck model used to derive (47) assumes a very small $Ni\alpha$, so a reduction in response is still observed at $Ni\beta \leqslant 2$. Also there is less reduction with $Ni\beta = 4$ for the case where $Ni\alpha = 4$ (Figure 8), relative to $Ni\alpha$ becoming infinitely

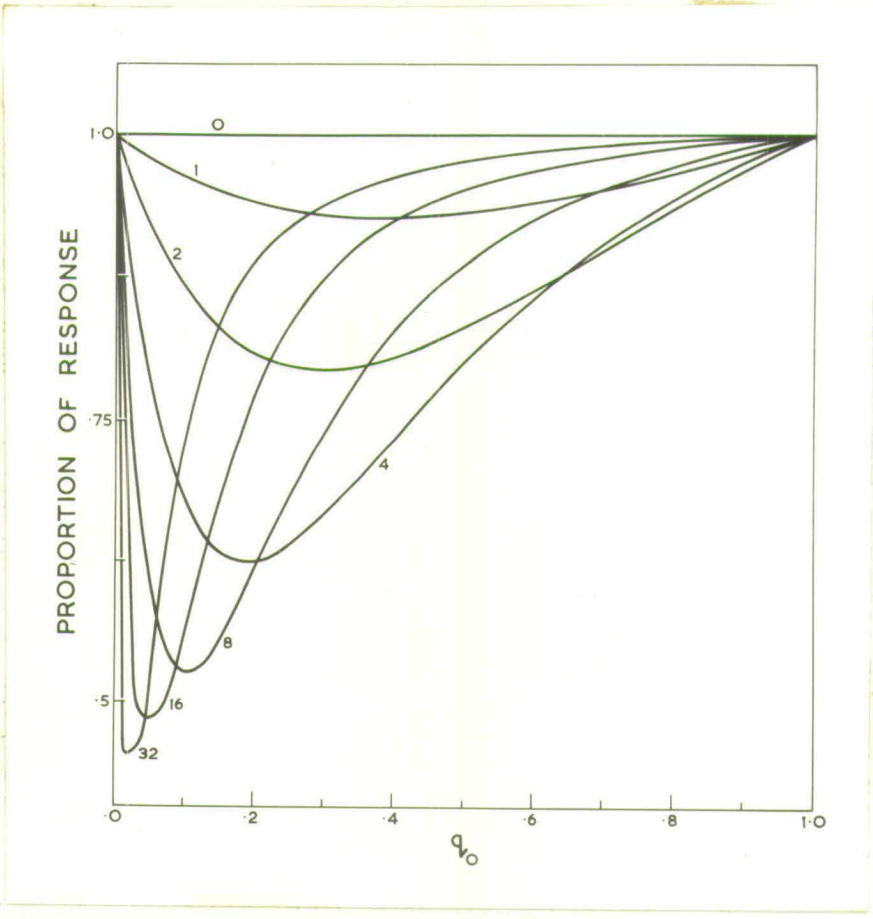


FIGURE 26 The influence of the initial frequency and value Niβ of a completely linked additive gene on the total response of an additive gene with very small effect. The response is measured as a proportion of that expected from a single gene with the same parameters.

small (Figure 26). These differences between the two models could be predicted from the size of effects, so it would appear that if a gene with sufficiently small $Ni\alpha$ was run by Monte Carlo simulation with a large number of replicates and with $Nc = 0$, a curve almost the same as Figure 26 could be drawn. Thus it would seem that the model of a bottleneck of effective population size used to derive equation (47) is an adequate description of the influence of an additive gene on the response of another completely linked additive gene with small $Ni\alpha$ value.

The data for Figure 26 is presented in an alternative form in Figure 27, in which the proportional reduction in response is plotted against $Ni\beta$ for a few values of $q_o$. Figure 27 is therefore analagous with Figures 12 and 13 for $Nc = 0$, but, as would be expected, the curves only resemble each other when $Ni\beta > Ni\alpha$, say.

Apart from being able to mimic the Monte Carlo results, the approach developed here has a useful predictive value for low $Ni\alpha$. If $Ni\alpha$ is small, Monte Carlo simulation is very inefficient since the response, $u(p_o)-p_o$, is small relative to its standard error. On the other hand, the bottleneck method only requires the inversion of three matrices, with no replication, to simulate the response for low $Ni\alpha$.

When the approach developed in this section for small $Ni\alpha$ was first introduced, it was argued in terms of a favourable allele B at a low initial frequency, yet with a chance of fixation

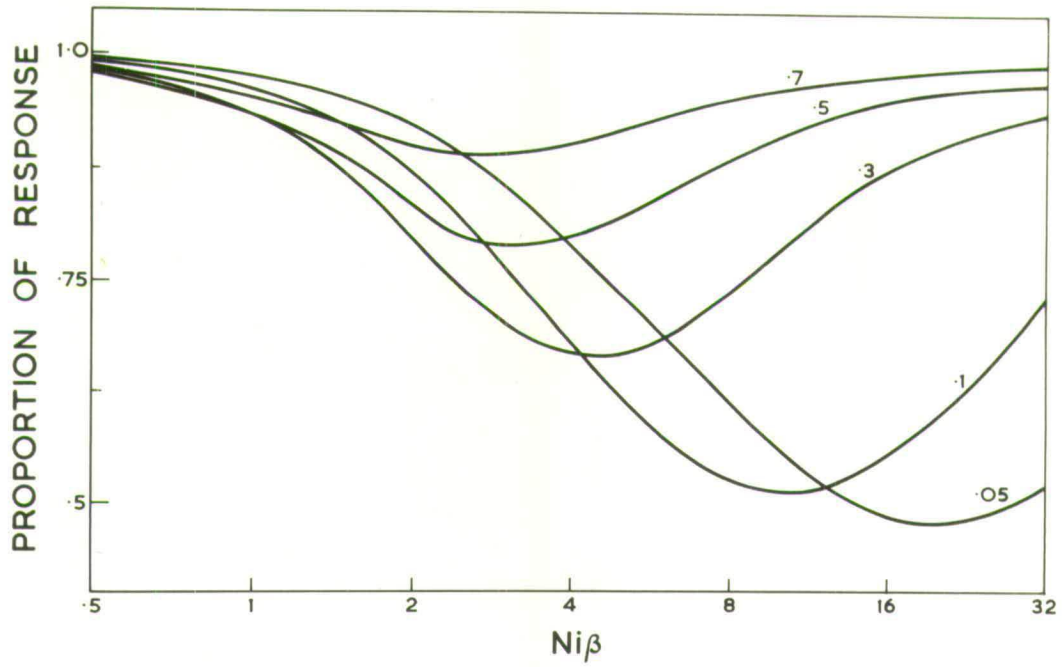


FIGURE 27 The influence of the value Niβ and initial frequency of a completely linked additive gene on the total response of an additive gene with very small effect. The response is measured as a proportion of that expected from a single gene with the same parameters.

close to unity because $Ni\beta$ was assumed to be large. As a result all A, a alleles would pass through a narrow bottleneck of effective population size for all response would be made among gametes containing B and these would initially be few in number. If B is lost when initially at low frequency, little reduction in the variance of A can occur since the number of b gametes in the population will always be close to 2N. However, it has been found that the maximum reduction in $u(p_o)$ occurs when $u(q_o) \sim 0.8$ for large $Ni\beta$, and, furthermore, genes of intermediate $Ni\beta$ values may reduce response more than those with larger $Ni\beta$, even if the initial frequency is low. An explanation for these results can easily be given. With the higher $Ni\beta$ values the response to selection of the B allele is more rapid, so that although there may be an initial bottleneck, the population of gametes having B, within which A is selected, rapidly expands. It can therefore be predicted that with large $Ni\beta$ values for the same, intermediate, $q_o$, more response in A will be made in later generations than with small $Ni\beta$. Such a result was observed in Figure 22 and in the half-lives shown in Table 7. Similarly, the maximum reduction in $u(p_o)$ does not occur when $u(q_o) = 1$. For the chance of fixation, $u(q_o)$, to approach one, it is necessary for $Ni\beta q_o$ to be greater than 2, approximately. Thus the bottleneck for A amongst gametes containing B is much less, in the early generations, than if a lower value of $q_o$ is found. It therefore turns out that the average reduction is larger with 80% of the population passing through a very narrow bottleneck and 20% being scarcely restricted

at all, than with all the populations of A passing through a larger bottleneck.

In the Monte Carlo simulation results it was found that the realised selection parameter, $\hat{N}i\alpha$ (the value of $Ni\alpha$ which would give the same chance of fixation for a single gene with the same initial frequency), is almost independent of the initial frequency, $p_o$, at least where $\alpha < \beta$. In the derivation of (47) it is actually assumed that $\hat{N}i\alpha/Ni\alpha$ is independent of $p_o$, and since $Ni\alpha$ is very small it is assumed that $\hat{N}i\alpha/Ni\alpha$ is also independent of $Ni\alpha$. This result is likely to hold fairly well for all $Ni\alpha < 1/2$, say, when changes in gene frequency are expected to be small, and the decline in the variance from drift remains closely proportional to $p_o(1 - p_o)$. It was found earlier, however, that $\hat{N}i\alpha/Ni\alpha$ may be much smaller for genes where $Ni\alpha$ is almost as large as $Ni\beta$ than for genes with smaller $Ni\alpha$ (Figures 14-17). The explanation seems to be that the bottleneck in population size of B gametes occurs in the early generations of selection and would thus be expected to influence most the response of genes A which would normally be responding rapidly at that time. Now genes with large $Ni\alpha$ make a high proportion of their response in the early generations, in other words they have a short half-life (Figure 20), and would therefore be particularly affected by a restriction in population size in the initial generations. Genes of smaller $Ni\alpha$ generally make most of their response after the bottleneck has been passed. Similarly, it was noted in Figures 14-17 that, particularly when B had a low initial frequency, there was more

reduction in $\hat{N}i\alpha/Ni\alpha$ for genes A of high than of low initial frequency. The same interpretation must hold, for genes A of high initial frequency have a shorter half-life than do genes of low frequency with the same $Ni\alpha$ (Figure 20), and so would be expected to be more affected by an early reduction in population size. This is likely to be most extreme when B has a low initial frequency and the initial bottleneck is very small. Turning back to the relation between $\hat{N}i\alpha/Ni\alpha$ and the magnitude of $Ni\alpha$, it was noted that it was not practicable by Monte Carlo simulation to find the limiting value of $\hat{N}i\alpha/Ni\alpha$ as $Ni\alpha \rightarrow 0$. However, the results obtained from (47) give a solution. For the example given in detail earlier, for which $q_0 = 0.1$, $Ni\beta = 32$ and $Nc = 0$, it was found that with $Ni\alpha = 32$, 16, 8, 4 and 2 the ratio $\hat{N}i\alpha/Ni\alpha$ was 0.26, 0.20, 0.18, 0.50 and 0.60, respectively. For very small $Ni\alpha$, the ratio for this example is seen to be $\hat{N}i\alpha/Ni\alpha = 0.73$ (Figure 26).

If B has a low initial frequency, the effective population size for A will be smaller among gametes having the B allele than among those with the b allele. Thus the expected response in $p_1$ will be smaller than in $p_2$, and negative disequilibrium will be observed both within lines (Figure 25) and between lines at the limit (Figure 18). However a necessary consequence of this hypothesis seems to be that positive disequilibrium would be expected if $q_0$ is greater than one-half. Evidence from the Monte Carlo simulations is not adequate on this point. In particular most of the runs in which

$q_o > 0.5$ have resulted in a chance of fixation $u(q_o) = 1$ and hence no disequilibrium at the limit. Positive disequilibrium raises problems on the consequences of recombination, for there is certainly no significant evidence from the Monte Carlo runs that low Nc values yield greater responses than do higher Nc values, other parameters remaining constant. Reductions in $u(p_o)$ are always small for $q_o > 0.5$ so this aspect will be difficult to study precisely; however some further investigation is clearly necessary.

## 8. DISCUSSION

The discussion in the previous sections of the thesis has been concerned with interpreting the Monte Carlo results obtained. Many of the assumptions in this Monte Carlo study and some of its limitations will now be outlined.

From the diffusion equation (section 3) it was argued that computer runs had only to be made at one level of population size. However the parameters $i\alpha$, $i\beta$ and c used were frequently much larger than those necessary for the diffusion approximation to hold, but it appears (Figures 10 and 24) that the conclusion that $Ni\alpha$, $Ni\beta$ and Nc are sufficient parameters is highly robust against departures from the underlying assumptions. Computer runs were usually made with as small a population size as possible in order to reduce computing time. In order that results would be more appropriate for populations of size larger than those simulated several assumptions were made in the selection procedure adopted. At the same time these approximations further reduced the amount of computation necessary. In particular, the algebra developed for infinite populations and used to simulate selection and recombination was entirely in terms of gametes, no distinction was made between the sexes and self fertilisation was permitted. The precision with which that process describes the real situation for a bisexual species must be largely a function of population size, the greater N the smaller any errors introduced by these approximations become. A similar type of inaccuracy was introduced into the definition of

the selective values (2), which are precise only for genes of small effect. Strictly, second and higher order terms in effects should have been included (e.g. $(i\alpha)^2$, $(i\beta)^2$) but then the results could not have been generalised in terms of $Ni\alpha$ and $Ni\beta$ to populations of different sizes.

The selective values, $i\alpha$ and $i\beta$, of the favourable alleles have been kept constant throughout the selection process, for which two important assumptions have been made. Firstly, the gene effects $\alpha$ and $\beta$ have been defined (5) as the difference in genotypic value between the homozygotes at a locus as a proportion of the phenotypic standard deviation, $\sigma$, so that for the selective values to remain constant, $\sigma$ itself must be unchanged. In making the same assumption, Robertson (1960) pointed out that although the genetic variance would be expected to decline during selection, at the same time the environmental component might increase as the level of homozygosity rises. Secondly, no account has been taken of natural selection which might be expected to alter the effective selective values of genes having correlated effects on fitness as gene frequencies move from their initial equilibrium.

The model of two additive loci each with two alternative alleles which has been studied is probably the simplest in which linkage could be included. In earlier Monte Carlo studies a larger number of additive loci have been simulated, but with the exception of Fraser and Hansche (1965) who were not concerned with selection limits, workers in this field have always used a model of equal effects and initial frequencies one-half at each locus (Fraser, 1957b;

Martin and Cockerham, 1960; Gill, 1963, 1965; Qureshi, 1963). However with the restricted model used here the initial frequencies, selective values and recombination fraction could be varied over a wide range and it was possible to obtain a fairly complete description of the selection limit for this model. In particular, the influence of one gene on the chance of fixation of the other could be studied in detail, and this approach seems more likely to lead to an understanding of the process than if only the change in the population mean, which is dependent on the responses of both genes, is considered. Thus any reduction in response can be viewed as a function of two components. Firstly, the bottleneck of population size within gametes containing the favourable gene B reduces the effective selection pressure ($\widehat{Ni\alpha}$) on the A gene, the reduction being approximately independent of the initial frequency of A. The effect of this bottleneck on the chance of fixation of A then depends on the slope of the curve of $u(p_o)$ against $\widehat{Ni\alpha}$. If only equal effects and initial frequencies are studied, it is unlikely that these essential parts of the process could be disentangled.

It is intended to continue this study to include more than two loci and non-additive gene effects, so at this stage there is little benefit in discussing these extensions in detail. The bottleneck model and the related model for equal effects outlined in the previous section are pertinent for all non-epistatic systems and the same general picture of reduction in response is likely to be found with dominant as well as additive genes. For the case where the affected gene has a very small $Ni\alpha$, it should be possible

to extend the matrix derivations to include the case of the influence of both a dominant gene on a linked additive gene, an additive on a dominant and finally a dominant on a dominant. Perhaps the most simple multi-locus model is where a chromosome has only one gene of large effect, and several genes of very small effect. Then the response of these smaller genes is likely to be affected only by the larger gene and the model reduces essentially to two loci. However when there is more than one locus of intermediate or large effect on the chromosome no definite conclusions can yet be given about their influence on a third gene.

This study has been further restricted to include only populations initially in linkage equilibrium. However Mather (1943) has argued that natural selection would favour balanced repulsion, but Wright (1952) has shown that selective values have to be large and linkage tight for much repulsion to be maintained. In general, if loci have no epistatic effects on fitness, an unselected closed random mating population would be expected to remain in equilibrium (Lewontin and Kojima, 1960). On the other hand, disequilibrium will almost certainly be found if the populations are derived from a cross between lines, whether unrelated or selected from the same original population, or from a cross of a highly selected line to the original stock. It is hoped to investigate these situations theoretically in some detail, for they have an important bearing on the problem of breaking selection limits. Some relevant information has come from this study, however.

It has been shown that with tight linkage there is usually negative disequilibrium ($\Delta_L$) in the chance of fixation of the gametes. Thus line crosses, or crosses of selected lines to a base population in equilibrium, will, on average, have an excess of repulsion heterozygotes. Response after the cross is therefore likely to be reduced when genes are tightly linked, and it may be necessary to relax selection for a few generations to allow recombination. Also, if gene effects are unequal, an initial period of reverse selection might be advantageous, so that the frequency of the smaller effect genes in repulsion gametes may be increased. Osman (1963) allowed varying periods of random mating before re-selecting crosses to the base population of a line of Drosophila melanogaster selected close to a limit for sternopleural bristle number. Although Osman concluded that, on average, the limit was reduced by this random mating, two of the four crosses making most response had undergone seven generations of relaxation before selection. Of course, a critical factor determining the effectiveness of any period of random mating is the population size which can be maintained during that time.

When there is complete linkage (Nc = 0) the new selection limit after crossing selected lines can be computed using Kimura's (1957) formula (equation 17), for there are only two alternative gametes in each possible cross. If the new limit is termed $v(p_o)$ for A, then, in general,

$$v(p_o) = [u(p_o)]^2 + 2[u(AB)u(aB) + u(Ab)u(ab)]\left[\frac{1 - e^{-Ni\alpha}}{1 - e^{-2Ni\alpha}}\right]$$

$$+ 2u(AB)u(ab)\left[\frac{1 - e^{-Ni(\alpha+\beta)}}{1 - e^{-2Ni(\alpha+\beta)}}\right] + 2u(Ab)u(aB)\left[\frac{1 - e^{-Ni(\alpha-\beta)}}{1 - e^{-2Ni(\alpha-\beta)}}\right] \quad (48)$$

where N and i are the population size and selection intensity during the period of re-selection. For one locus, Robertson (1960) found that if the further selection from the cross has twice the $Ni\alpha$ value as has the original selection, then the expected limit, $v(p_o)$, would be the same as for selecting one line from the original population with twice the original value of $Ni\alpha$. It appears that the same result holds approximately for linked loci. An example is given in Table 10 where in the original sub-lines $Ni\alpha = 4$ and $Ni\beta = 8$. Data is taken from Figures 3-7 and limits for the cross were calculated by equation (48).

TABLE 10 Total response $(v(p_o)-p_o)$ from selecting in two sub-lines and re-selecting their cross as a proportion of the total response from selecting in one line. In the sub-lines $Ni\alpha = 4$ and $Ni\beta = 8$, otherwise $Ni\alpha = 8$ and $Ni\beta = 16$. $Nc = 0$.

| $p_o$ \ $q_o$ | .05 | .1 | .3 | .5 | .7 |
|---|---|---|---|---|---|
| .05 | 0.73 | 1.29 | 1.22 | 1.26 | 0.95 |
| .1 | 1.02 | 1.21 | 1.15 | 1.02 | 1.03 |
| .3 | 1.19 | 1.02 | 1.06 | 1.01 | 1.01 |
| .5 | 0.87 | 1.07 | 0.98 | 1.00 | 0.99 |
| .7 | 1.00 | 1.11 | 1.00 | 1.00 | 1.00 |

In Table 10 it appears that, while on average rather more response is made by splitting the original population and re-selecting

the cross, the differences between the methods are never large. Further investigation is clearly necessary to determine whether these alternatives ever differ much in efficiency.

Returning to the problem of one cycle of selection from the base population, it was found that, with an additive model, if linkage influenced response it was always in the direction of a reduction in the limit. What evidence is there from animal populations that this occurs? Robertson (1965) selected 10 parents out of 25 scored for each sex for low sternopleural bristle number in Drosophila melanogaster. In five lines crossing over was suppressed on chromosomes 2 and 3, in another five lines crossing over was permitted. The base population had a mean of about 17.8, and the averages of the five line sets for bristle number were as follows:

| Generation | 5 | 10 | 14 |
|---|---|---|---|
| Suppressed | 15.7 | 14.8 | 14.5 |
| Unsuppressed | 15.4 | 13.9 | 13.1 |

Also, after 13 generations, every line in which crossing over was permitted had responded more than in every line in which it was suppressed. The pattern of response in these lines is clearly similar to the pattern observed in the Monte Carlo runs. In the early generations the response is about the same, whether recombination occurs or not, but with tight linkage the response rate slows down much more rapidly in the later generations.

It is important to draw attention at this point to the degree of linkage simulated in the Monte Carlo runs. Generally the largest Nc value used was Nc = 1, which with 20 selected parents

implies a recombination fraction of only 0.05. It has been found that, at least for two loci, greatest reductions in the change in the population mean are found when the gene effects are approximately equal (Figure 19). In this case, there appears to be an almost linear regression of response against Nc, transformed to a scale of $1/(2Nc + 1)$. The example of Figure 19 shows that when $p_o = q_o = 0.1$ and $Ni\alpha = 8$, the largest reductions occur with $Ni\beta = 8$ when about 70% of the response is made with $Nc = 0$ relative to $Nc = \infty$. Using the $1/(2Nc + 1)$ transformation, the expected responses for this example with $N = 20$ would be, as a proportion of the response with $Nc = \infty$,

98.6, 97.3, 94.0, 90.0, 85.0, 78.6 and 70.0%

for c = .5, .25, .1, .05, .025, .01 and .0 respectively.

These results illustrate a general conclusion that can be gained from this study, for only when widely different recombination fractions are compared is much difference in response to be expected. Thus in the above example the greatest difference in response observed for a doubling or halving of the recombination fraction is only 5%. This occurs in the range around $c = 1/2N$, where the curve of $1/(2Nc + 1)$ against Nc has greatest slope (Figure 2).

These results have a bearing on the optimum intensity of artificial selection which should be applied in order to maximise the selection limit. In many selection programmes the number of progeny (T) that can be recorded is fixed. With one locus, or no linkage, the optimum proportion (N/T) to select is therefore the value of N/T which maximises Ni, which turns out to be one-half (Dempster, 1955; Robertson, 1960). Also, Ni is the same whether

N/T or (T - N)/T individuals are selected, for any value of N and T. When there are linked loci, it can be expected that the optimum intensity of selection will be rather lower, for if more than half the individuals are selected, although $Ni\alpha$ is reduced relative to N/T = 0.5, at the same time Nc is increased and $Ni\beta$ reduced. Both the latter would generally increase the response for A. Further, the limit will no longer be symmetric about N/T = 0.5, very intense selection being less successful than very weak selection in the long run. An example is given in Table 11 for the case of T = 40, $\alpha = \beta = 0.5$, c = 0.025 and $p_o = q_o = 0.1$. It is assumed that the response is proportional to $1/(2Nc + 1)$ for given $Ni\alpha$ and $Ni\beta$. The results are obtained from interpolation of Monte Carlo data and are approximate.

TABLE 11 Chance of fixation of an additive gene when 40 individuals (T) are recorded, $\alpha = \beta = 0.5$ and $p_o = q_o = 0.1$.

| | Proportion selected (N/T) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | .05 | .1 | .25 | .4 | .5 | .6 | .75 | .9 | .95 |
| No linkage | .34 | .51 | .71 | .78 | .80 | .78 | .71 | .51 | .34 |
| c = 0.025 | .31 | .46 | .61 | .66 | .70 | .70 | .65 | .49 | .33 |
| c = 0 | .30 | .45 | .52 | .60 | .61 | .60 | .52 | .45 | .30 |

When N/T = 0.5, Nc = 0.5 for c = 0.025 so the above example is relatively sensitive to changes in Nc as the proportion of the population selected is altered from one-half. However it can be seen that the optimum is still close to N/T = 0.5 when c = 0.025 and the

curve of $u(p_o)$ against N/T is not very skewed. Of course, with no recombination the optimum remains at N/T = 0.5 and the curve is symmetric. Thus when designing selection programmes it would appear that considerations of linkage should not influence greatly the intensity of selection to be practised. However more drastic effects might be found with more than two loci.

## 9. SUMMARY

A theoretical investigation was made of the influence of linkage on limits to artificial selection in small populations. Most results were obtained by Monte Carlo simulation.

A model of two additive loci, each with two alleles, was used. The difference between the effects of the two homozygotes was expressed as a proportion of the phenotypic standard deviation and defined as $\alpha$ and $\beta$ for the loci with favourable alleles A and B, respectively. These alleles had initial frequency $p_o$ and $q_o$, respectively. It was assumed that the base population was in linkage equilibrium, and that the recombination fraction, c, was the same for both sexes.

It was shown that, if the effective population size is N and the selection differential is i standard deviations, the selection limit is a function of only $p_o$, $q_o$, $Ni\alpha$, $Ni\beta$ and Nc, and the time scale of the selection process is proportional to N. Thus it was necessary for Monte Carlo computer runs to be made with only one population size.

The chance of fixation (the expected gene frequency at the limit) of A, $u(p_o)$, may be greatly reduced if the loci are tightly linked and if $\beta$ is not less than about one-half of $\alpha$. The chance of fixation is never increased by linkage if the population is initially in equilibrium.

Unless $\alpha$ and $\beta$ differ widely, the decline in $u(p_o)$ with reduction in Nc is approximately linearly related to the rate of

breakdown of linkage disequilibrium in small populations, which is proportional to $1/(2Nc + 1)$.

The larger $Ni\beta$, the greater the maximum reduction in $u(p_o)$ and the lower the initial frequency of B at which this maximum occurs. For large $Ni\beta$, B has most influence if $q_o \sim 0.8/Ni\beta$. If $q_o$ is higher than this, a gene, B, with smaller $Ni\beta$ may have a larger effect on $u(p_o)$.

It was shown that the influence of the linked gene, B, is approximately independent of the initial frequency of A.

The total response of the population mean, a function of both the responses and effects at each locus, is most influenced by tight linkage when the loci have approximately equal effect and the favourable alleles have a low initial frequency.

In the early generations the rate of selection advance is not affected by linkage, but in later generations the rate may become very much slower than with free recombination.

The degree of linkage has little influence on the optimum intensity of artificial selection.

Theoretical models were developed to interpret these results. If the initial gene frequencies are low, the favourable coupling gamete, AB, is rare, and in a small population AB may never be formed if linkage is tight. If B has a high chance of fixation and $\alpha < \beta/2$, selection for A can be viewed as selection in a population whose effective size is the number of gametes that contain the B allele. If B has a low initial frequency this effective population size is initially very small and the change in gene frequency of A is therefore reduced.

The model was developed to give a method for simulating the case of complete linkage, with very small $N i \alpha$, which does not require the use of Monte Carlo techniques.

ACKNOWLEDGEMENTS

I am particularly indebted to Dr. Alan Robertson, O.B.E., F.R.S. for many helpful discussions and ideas during the course of this work.

I am also very pleased to thank :

The Agricultural Research Council for a Postgraduate Studentship,

Professor C.H. Waddington, C.B.E., F.R.S. for laboratory facilities in the Institute of Animal Genetics,

The Directors and staffs of the Edinburgh University Computer Unit, the Manchester University Computer Unit and the Atlas Computer Laboratory, S.R.C., Chilton for computational facilities,

Miss Bettie McCreesh for technical assistance, Mr. E.D. Roberts for drawing the figures and Miss Doris Manning for typing the manuscript, and Mr. J.H. Louw for much stimulating discussion throughout this study.

## REFERENCES

ALLAN, J.S. and ROBERTSON, A. 1964. The effect of an initial reverse selection upon total selection response. Genet. Res. 5: 68-79.

BENNETT, H.J. 1954. On the theory of random mating. Ann. Eugenics 18: 311-317.

DEMPSTER, E.R. 1955. Genetic models in relation to animal breeding problems. Biometrics 11: 535-536.

EWENS, W.J. 1963. Numerical results and diffusion approximations in a genetic process. Biometrika 50: 241-249.

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

FINKBEINER, D.T. 1960. Introduction to matrices and linear transformations. Freeman, San Francisco.

FRASER, A.S. 1957a. Simulation of genetic systems by automatic digital computers. I. Introduction. Aust. J. Biol. Sci. 10: 484-491.

FRASER, A.S. 1957b. Simulation of genetic systems by automatic digital computers. II. The effects of linkage on the rates of advance under selection. Aust. J. Biol. Sci. 10: 492-499.

FRASER, A.S. and HANSCHE, P.E. 1965. Simulation of genetic systems. Major and minor loci. Genetics Today, Ed. S.J. Geerts. Pergamon Press. pp.507-516.

GEIRINGER, H. 1944. On the probability theory of linkage in Mendelian heredity. Ann. Math. Stat. 15: 25-57.

GILL, J.L. 1963. Effect of population size, selection intensity, linkage and non-additive variability upon genetic change in simulated populations. Ph.D. Thesis, Iowa State University.

GILL, J.L. 1965. Effects of finite size on selection advance in simulated genetic populations. Aust. J. Biol. Sci. 18: 599-617.

GRIFFING, B. 1960. Theoretical consequences of truncation selection based on the individual phenotype. Aust. J. Biol. Sci. 13: 307-343.

KENDALL, M.G. 1943. The advanced theory of statistics. Volume 1. Griffing, London.

KIMURA, M. 1955. Stochastic processes and distribution of gene frequencies under natural selection. Cold Spr. Harb. Symp. quant. Biol. 20: 33-55.

KIMURA, M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28: 882-901.

KIMURA, M. 1958. On the change of population fitness by natural selection. Heredity 12: 145-167.

KIMURA, M. 1963. A probability method for treating inbreeding systems, especially with linked genes. Biometrics 19: 1-17.

LATTER, B.D.H. 1965. The response to artificial selection due to autosomal genes of large effect. I. Changes in gene frequency at an additive locus. Aust. J. Biol. Sci. 18: 585-598.

LEWONTIN, R.C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49: 49-67.

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

MARTIN, F.G. and COCKERHAM, C.C. 1960. High speed selection studies. Biometrical Genetics, Ed. O. Kempthorne, Pergamon Press. pp. 35-45.

MATHER, K. 1943. Polygenic inheritance and natural selection. Biol. Rev. 18: 32-64.

NEI, M. 1963. Effect of selection on the components of genetic variance. Statistical Genetics and Plant Breeding, Ed. W.D. Hanson and H.F. Robinson. Publ. 982, National Academy of Sciences, National Research Council, Washington, D.C. pp. 501-515.

OSMAN MOUSA, H.E. 1963. Overcoming selection limits by the introduction of new genetic material from unselected and randomly mating populations. Ph.D. Thesis, Edinburgh University.

QURESHI, A.W. 1963. A Monte Carlo evaluation of the role of finite population size and linkage in response to continuous mass selection. Technical Report MC 6, Statistical Laboratory, Iowa State University.

ROBERTSON, A. 1960. A theory of limits in artificial selection. Proc. Roy. Soc. (Lond) 153B: 234-249.

ROBERTSON, A. 1965. Selection in plants and animals. Proceedings of Mendelian Symposium, London. Royal Society and Genetical Society. In press.

WRIGHT, S. 1933. Inbreeding and recombination. Proc. Nat. Acad. Sci., Wash. 19: 420-433.

WRIGHT, S. 1952. The genetics of quantitative variability. Quantitative Inheritance, Ed. E.C.R. Reeve and C.H. Waddington. H.M.S.O., London. pp. 5-41.

# APPENDIX A

## SOME OBSERVATIONS ON ASYMMETRICAL CORRELATED RESPONSES TO SELECTION

B. B. Bohren[1], W.G. Hill[2] and A. Robertson,

Institute of Animal Genetics, Edinburgh 9

[1] Purdue University Agricultural Experiment Station, Lafayette, Indiana, U.S.A. This study supported by the U.S. Educational Commission in the United Kingdom.

[2] Agricultural Research Council postgraduate student.

Artificial selection applied to one character almost always leads to changes in others. The theory of such "correlated responses" is well known and has recently been reviewed by Falconer (1960a). In this, the genetic correlation between the two characters plays an important part and determines the predicted pattern of the correlated responses found in different experiments, e.g., the response in character 2 on selection for character 1 compared to that in 1 on selection for character 2 or the comparison of the responses in 2 on selection for 1 in opposite directions. Any discordance of the pattern of correlated responses from expectation will be termed an "asymmetrical correlated response". The same measurement made under two different environments can be considered as two separate "characters".

Falconer (1960b) selected mice for growth rate on high and low planes of nutrition and observed the correlated responses on the alternate nutritional level. The realized genetic correlations were equal for the first four generations of selection (0.67, 0.65) but were markedly different for generations 5 to 13 (1.25, -0.02). The asymmetry was attributed to changes in the basic parameters due to the selection applied, and large changes in the phenotypic standard deviations were observed. Asymmetry of the realized genetic correlations was also observed by Bell and McNary (1963) who selected Tribolium castaneum for increased pupal weight in both a wet and a dry environment, and by Yamada and Bell (1963) where selection was for

increased and decreased 13 day larval weight in Tribolium castaneum under good and poor nutritional levels.

Similar results have been observed in poultry by Siegel (1962) and Nordskog and Festing (1962). The former selected for four generations for body weight and breast angle, and found a realized genetic correlation of about .55 when selection was for body weight and a value of about .45 when selection was for breast angle. The latter workers selected in both high and low directions for body weight and egg weight, and observed asymmetry of the realized genetic correlations between body and egg weights when either the direction of selection or the trait being selected was considered. In both of these papers, the asymmetry was attributed to differing genetic variances or heritabilities for the two traits.

Clayton, Knight, Morris and Robertson (1957) observed asymmetry in response of sternopleural bristle number in Drosophila melanogaster to selection for increased and decreased sternital bristle number. The results were somewhat erratic, which led the authors to conclude that gene drift may play an important part in the correlated response when the genetic correlation is low. In general, however, there was a positive correlated response when selection was for increased sternital bristle number and no correlated response when selection was for low sternital bristle number.

On the other hand, some selection experiments show a close fit of expected to observed correlated responses. For example, Reeve and Robertson (1953) selected for wing and thorax lengths in Drosophila

melanogaster and found good agreement between estimates of the genetic correlation between the two characters in the base population and the realized genetic correlations in the populations selected for each trait separately.

The frequency with which asymmetrical correlated responses have been found suggests that some mechanism other than genetic sampling is affecting the correlated response in these populations. The purpose of this study was to re-examine the theory of correlated response and if possible to establish the conditions in which asymmetry of correlated response to selection was to be expected.

## The Model

It has been shown by Falconer (1960a) that the correlated response in trait 2 from selection for trait 1 would be

$$C\,R_{2.1} = \bar{i}_1\, h_1\, h_2\, r_G\, \sigma_2 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where $\bar{i}_1$ is the selection intensity for trait 1 in standard units, $h_1$ and $h_2$ are the square roots of the heritabilities for traits 1 and 2, respectively, $r_G$ is the genetic correlation between the two traits and $\sigma_2$ is the phenotypic standard deviation in trait 2. Dividing both sides by $\bar{i}_1\sigma_2$ results in a standardized correlated response ($CR'_{2.1}$) or the correlated response in standard deviations in trait 2 for each standard deviation of selection in trait 1. Thus,

$$\frac{CR_{2.1}}{\bar{i}_1\sigma_2} = CR'_{2.1} = h_1\, h_2\, r_G \quad \ldots\ldots\ldots\ldots \quad (2)$$

In a similar manner, $\frac{CR_{1.2}}{i_2\sigma_1} = CR'_{1.2}$ can be obtained, and it is seen

that

$$CR'_{2.1} = CR'_{1.2} = h_1 h_2 r_G = \text{Cov } G/\sigma_1 \sigma_2 \ldots\ldots (3)$$

The standardized correlated response should be the same in the first generation whether selection is on trait 1 or on trait 2 or whether the selection is upwards or downwards.

When the correlated response is measured over several generations, selection may change the value of the parameters themselves in such a way that the standardized responses, as measured in the two different populations, are asymmetrical and different from those predicted on the basis of the original parameters. This follows the suggestion of Falconer (1960b) that the asymmetrical responses he observed were the consequence of changes in parameters due to selection. Large changes in the phenotypic standard deviations were observed, and the potential effect of these changes on the standardized correlated response is evident from equation (3).

The three parameters of interest in (3) are the genetic covariance and the phenotypic standard deviations, and consideration is centered on how these parameters can change during selection for each trait involved.

The genetic covariance between two traits, as calculated in any population by the usual analysis of covariance technique, can be caused either by linkage disequilibrium of genes affecting the two traits independently or by the pleiotropic effects of single genes. In the case of linkage, the population would tend toward equilibrium at variable rates depending upon the cross over distance between the genes. The effect of linkage on the correlated response would be similar to that of pleiotropic genes, except that, as crossing over occurred

and the population approached equilibrium, the effect of linked genes on the asymmetry of correlated response over a number of generations would be less than that of pleiotropic genes. Therefore, only pleiotropic genes are considered, as the most extreme and constant case.

A genetic model for correlated responses was then constructed, and the expected values of the parameters and the correlated responses obtained from a Sirius computer for each of 9 generations of selection. In the first series of selections, only additive gene effects were considered since these would appear to be least likely to yield asymmetrical correlated responses. Four types of loci are considered in the model, with the following effects of a gene substitution on the two traits:

| | (A) | (B) | (C) | (D) |
|---|---|---|---|---|
| Trait 1 | $\alpha$ | $\beta_1$ | $\gamma_1$ | 0 |
| Trait 2 | 0 | $\beta_2$ | $-\gamma_2$ | $\delta$ |

Loci A and D affect the two traits independently. Loci B and C affect both traits, the former making a positive and the latter a negative contribution to the covariance. The substitution effects shown refer to one-half the difference between the alternative homozygotes. Only one locus of each type with additive effects is assumed. The existence of more than one locus having the same type of correlated effects would not affect the occurrence of asymmetry, but only the rate and pattern of its development, as we shall see later. It is assumed that the frequencies of the genes at each locus are $q_A$, $q_B$, $q_C$ and $q_D$, respectively, the first three referring to the allele with positive effect on trait 1. For this model it is seen that

$$\text{Cov } G = 2q_B (1-q_B) \beta_1 \beta_2 - 2q_C (1-q_C) \gamma_1 \gamma_2,$$

$$V_1 = 2q_A (1-q_A) \alpha^2 + 2q_B (1-q_B) \beta_1^2 + 2q_C (1-q_C) \gamma_1^2 + V_{E1},$$

$$V_2 = 2q_B (1-q_B) \beta_2^2 + 2q_C (1-q_C) \gamma_2^2 + 2q_D (1-q_D) \delta^2 + V_{E2},$$

where $V_{E1}$ and $V_{E2}$ are the environmental variances of the two traits.

The computer was programmed to obtain the expected gene frequency at each locus for each generation. For example, the change in gene frequency at the A locus due to selection for trait 1 is $\frac{\bar{i}_1 \alpha}{\sigma_1} q_A (1-q_A)$, $\bar{i}_1 \alpha/\sigma_1$ being the selective advantage of the gene at that locus (Griffing, 1960). The new gene frequencies were used to calculate the genetic covariance, the genetic and phenotypic variances, the mean of each trait and the standardized correlated response for each generation when selection was on either trait 1 or trait 2. In all models $\bar{i}_1 = \bar{i}_2 = 1.0$, corresponding to a retention of 40% of the individuals as parents, except for models (ii) and (iv), in which $\bar{i}_1 = \bar{i}_2 = 0.5$, corresponding to 70% retention. In all runs the environmental variance was arbitrarily set equal to the genetic variance when all gene frequencies were one-half. The initial heritabilities of traits 1 and 2 were then close to one-half in all models. Because the above formula for the change in gene frequency was used for selection on the two characters, the correlated response is always symmetrical in the first generation. This formula does not in fact hold for genes with large effects and such genes could well produce asymmetry in the very first generation of selection. This appears to be most important under conditions when the gene selected for is at frequencies greater than 0.8 and when $\bar{i}\alpha/\sigma_1$ is greater than unity.

The asymmetry of the genetic covariances generated by the first generation of selection may be expressed algebraically as the difference between the genetic covariances after one generation of selection on trait 1 and trait 2, respectively. At the B locus, for instance, the genetic covariance is $2\beta_1\beta_2 q_B(1-q_B)$. If the gene frequency is increased by $q_B$, then the covariance is increased by

$$2\beta_1\beta_2 \left[ (1-2q_B)\,\Delta q_B - (\Delta q_B)^2 \right] .$$

Inserting the expressions for the change in gene frequency on selection for the two characters into this, and including the C locus, we obtain for the difference in covariance,

$$\text{Cov } G_1 - \text{Cov } G_2 =$$

$$2\beta_1\beta_2 q_B(1-q_B) \left[ (1-2q_B)\left( \frac{\bar{I}_1\beta_1}{\sigma_1} - \frac{\bar{I}_2\beta_2}{\sigma_2} \right) - q_B(1-q_B) \left( \frac{\bar{I}_1^2\beta_1^2}{V_1} - \frac{\bar{I}_2^2\beta_2^2}{V_2} \right) \right]$$

$$-2\gamma_1\gamma_2 q_C(1-q_C) \left[ (1-2q_C)\left( \frac{\bar{I}_1\gamma_1}{\sigma_1} + \frac{\bar{I}_2\gamma_2}{\sigma_2} \right) - q_C(1-q_C) \left( \frac{\bar{I}_1^2\gamma_1^2}{V_1} - \frac{\bar{I}_2^2\gamma_2^2}{V_2} \right) \right] \ldots\ldots (4) .$$

This equation is generalized to n loci affecting the two traits as

$$\text{Cov } G_1 - \text{Cov } G_2 =$$

$$2\sum_{k=1}^{n} \lambda_k \mu_k q_k(1-q_k) \left[ (1-2q_k)\left( \frac{\bar{I}_1\lambda_k}{\sigma_1} - \frac{\bar{I}_2\mu_k}{\sigma_2} \right) - q_k(1-q_k)\left( \frac{\bar{I}_1^2\lambda_k^2}{V_1} - \frac{\bar{I}_2^2\mu_k^2}{V_2} \right) \right] \ldots (5)$$

where $q_k$ is the frequency of one allele at the kth locus and $\lambda_k$ and $\mu_k$ are one-half the homozygote differences in traits 1 and 2 respectively, and can have either positive or negative values.

Expressions for the change in $V_1$ and $V_2$ on selection for the two characters can be obtained by substituting $\lambda_k \mu_k$ outside the square brackets in (5) by $\lambda_k^2$ and $\mu_k^2$ respectively. It then follows that a

symmetrical change in the covariance will also mean symmetry in the contributions of the B and C loci to the variance of the two characters.

Equation (4) consists of four terms in two pairs. Inside the square brackets, two terms have linear gene effects in them, and will both be zero when the gene frequency is 0.5. At usual selection intensities, and from what is known of gene effects, it is unlikely that expressions like $\bar{i}_1\beta_1/\sigma_1$ will be greater than 0.5. The second pair of terms contain such expressions squared and will therefore be smaller than the first pair. The gene frequency component in these will be at a maximum when the gene frequency is 0.5.

Of the components containing gene effects, it will be seen that three contain differences and only one contains a sum. From this, it is established that the simplest condition for asymmetry is the presence of C type loci making a negative contribution to the genetic covariance, with frequencies other than 0.5. From the entire gene frequency expression entering into this term, $q(1-q)(1-2q)$, the greatest absolute contribution to the asymmetry in the covariance will occur at frequencies of 0.2 or 0.8.

It is hardly surprising that the effects of A and D type loci do not appear directly in (4). They are of course involved in $\sigma_1$ and $\sigma_2$. It then follows that changes in the frequency of alleles at these loci are not of great importance and are exactly equivalent to changes in the environmental variance of either of the characters or to changes in the gene effects at these loci. This was confirmed by the computer results.

The other three terms in (4) contain differences in gene effects. They are more accurately differences in the selective advantage of the

genes under the two kinds of selection. The two containing square terms in gene effects will have maximum effects at gene frequencies around 0.5, while the other will have a maximum at 0.2 or 0.8. But, if gene effects are small, the term containing $(1-2q_B)$ will be dominating in the early generations.

Equation (4) gives the expected change in one generation. In t generations, the two linear terms will be multiplied approximately by t, but the squared terms approximately by $t^2$. In situations in which changes in the covariance in the first generation are entirely due to squared terms, the asymmetry in the covariance will then increase as the square of the number of generations. An example will be given among the computer results.

The Computer Results

The computer results shown in Tables 1 and 2 give the standardized direct and correlated responses accumulated over 9 generations of selection of the same intensity for the two traits. Various combinations of gene effects and initial frequencies have been chosen to exemplify the conclusions drawn from (4). The essential features of the gene effects chosen in the different models are given below. The comparison of standard deviations has been calculated for all gene frequencies at 0.5.

(i), (ii) and (iii). The B and C locus effects are the same - $\sigma_1 = \sigma_2$ for (i) and (ii) but $\sigma_1 < \sigma_2$ for (iii)

(iv) No variation at the C locus, $\beta_1 = \beta_2$ but $\sigma_1 < \sigma_2$.

(v) B locus effects twice those at C, and $\sigma_1 = \sigma_2$.

Table 1. Standardized direct responses (R) and standardized correlated responses (CR) after nine generations of selection on trait 1 (T1) and trait 2 (T2), with $h_1^2 \sim h_2^2$.

| Initial gene frequencies $q_A, q_B, q_C, q_D$ | | Gene effects T1 $\alpha, \beta_1, \gamma_1$ T2 $\beta_2, \gamma_2, \delta$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (i) 1, 1, 1 / 1, 1, 1 | | (ii) 0, 1, 1 / 1, 1, 0 | | (iii) 1, 1, 1 / 1, 1, 2 | | (iv) 0, 1, 0 / 1, 0, 1 | | (v) 1, 2, 1 / 2, 1, 1 | | (vi) 1, 1, 2 / 1, 2, 1 | | (vii) 1, 2, 1 / 1, 2, 1 | |
| | | T1 | T2 | T1 | T2 | T1 | T2 | T1 | T2 | T1 | T2 | T1 | T2 | T1 | T2 |
| (a) .5, .5, .5, .5 | R | 1.87 | 1.87 | 2.99 | 2.99 | 1.87 | 1.79 | 2.17 | 2.99 | 1.79 | 1.79 | 1.79 | 1.79 | 1.79 | 1.79 |
| | CR | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.47 | 2.11 | -0.61 | -0.61 | -0.39 | -0.39 | -0.46 | 0.46 |
| (b) .5, .2, .5, .5 | R | 2.28 | 2.28 | 3.61 | 3.61 | 2.28 | 2.10 | 3.38 | 3.61 | 2.29 | 2.29 | 2.10 | 2.10 | 2.29 | 2.10 |
| | CR | 0.39 | 0.39 | 0.72 | 0.72 | 0.26 | 0.36 | 2.30 | 2.93 | 0.93 | 0.93 | -0.11 | -0.11 | -0.19 | 0.94 |
| (c) .5, .5, .2, .5 | R | 2.28 | 1.55 | 3.61 | 2.23 | 2.28 | 1.54 | 2.17 | 2.99 | 2.10 | 1.54 | 2.29 | 1.35 | 2.10 | 1.35 |
| | CR | -0.39 | 0.38 | -0.71 | 0.97 | -0.26 | 0.38 | 1.46 | 2.11 | 0.11 | 0.69 | -0.93 | 0.13 | -0.94 | 0.72 |
| (d) .5, .5, .5, .2 | R | 1.87 | 2.28 | 2.99 | 2.99 | 1.87 | 2.29 | 2.17 | 3.61 | 1.79 | 2.10 | 1.79 | 2.10 | 1.79 | 2.10 |
| | CR | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.54 | 2.09 | 0.40 | 0.39 | -0.40 | -0.39 | -0.47 | 0.46 |
| (e) .5, .2, .2, .5 | R | 2.64 | 1.97 | 4.04 | 2.97 | 2.64 | 1.86 | 3.37 | 3.61 | 2.56 | 2.06 | 2.56 | 1.66 | 2.56 | 1.66 |
| | CR | 0.00 | 0.77 | 0.00 | 1.72 | 0.00 | 0.74 | 2.30 | 2.93 | 0.66 | 1.22 | -0.66 | 0.41 | -0.67 | 1.21 |
| (f) .5, .8, .3, .5 | R | 1.81 | 1.31 | 2.77 | 1.65 | 1.81 | 1.26 | 0.95 | 2.23 | 1.55 | 1.16 | 1.87 | 1.23 | 1.55 | 1.23 |
| | CR | -0.64 | -0.13 | -1.54 | -0.33 | -0.43 | -0.13 | 0.61 | 0.95 | -0.32 | 0.07 | -1.04 | -0.35 | -1.06 | 0.11 |
| (g) .5, .3, .8, .5 | R | 1.81 | 2.51 | 2.77 | 3.95 | 1.81 | 2.28 | 2.98 | 3.45 | 1.87 | 2.40 | 1.55 | 2.46 | 1.87 | 2.46 |
| | CR | 0.64 | -0.13 | 1.54 | -0.15 | 0.43 | -0.11 | 2.03 | 2.75 | 1.04 | 0.47 | 0.32 | -0.74 | 0.27 | 0.53 |

Table 2. Standardized direct responses (R) and standardized correlated responses (CR) after nine generations of selection on trait 1 (T1) and trait 2 (T2), with $h_1^2 \sim 2h_2^2$.

| Initial gene frequencies $q_A, q_B, q_C, q_D$ | | Gene effects T1 $\alpha, \beta_1, \gamma_1$ T2 $\beta_2, \gamma_2, \delta$ | | | |
|---|---|---|---|---|---|
| | | (i) 1, 1, 1 1, 1, 1 | | (v) 1, 2, 1 2, 1, 1 | |
| | | T1 | T2 | T1 | T2 |
| (a) .5, .5, .5, .5 | R | 1.87 | 1.25 | 1.79 | 1.15 |
| | CR | 0.00 | 0.00 | 0.29 | 0.44 |
| (e) .5, .2, .2, .5 | R | 2.64 | 1.24 | 2.56 | 1.35 |
| | CR | 0.00 | 0.69 | 0.47 | 1.23 |

(vi) B locus effects one half those at C, and $\sigma_1 = \sigma_2$.

(vii) The B locus has the greater effect on trait 1 but the C locus on trait 2, and $\sigma_1 = \sigma_2$.

In all models, the environmental variance for both characters was set equal to the genetic variance when all gene frequencies were 0.5.

The main points of interest in the correlated responses are as follows, classified according to the gene frequency combinations:

(a) All gene frequencies equal to 0.5. Only the squared terms in (4) can then contribute to the change in covariance. There is symmetry for all effect models except (iv) ($\beta_1/\sigma_1 \neq \beta_2/\sigma_2$) and (vii). In the latter, the selection for trait 1 causes most change in gene frequency away from 0.5 in a B locus and therefore reduces the genetic covariance. But selection for trait 2 changes most the frequency at a C locus and therefore increases the covariance. The asymmetry in the covariance increases as $t^2$ in the early generations (Fig.1). In (iii) the two squared terms are not zero but cancel out.

(b) $q_B \neq 0.5$. There is now slight asymmetry in (iii), arising from the linear term since, though $\beta_1 = \beta_2$, $\sigma_1 \neq \sigma_2$.

(c) $q_C \neq 0.5$. This is the situation to which attention was drawn earlier of a C locus with a frequency away from 0.5, which will lead to an asymmetry of covariance increasing linearly with time in all situations. The actual response when all gene effects are equal is shown in Figure 2.

(d) $q_D \neq 0.5$. In addition to models (iv) and (vii) a trivial asymmetry in correlated response is now found in models (v) and (vi) because asymmetry has developed in $\sigma_2$.

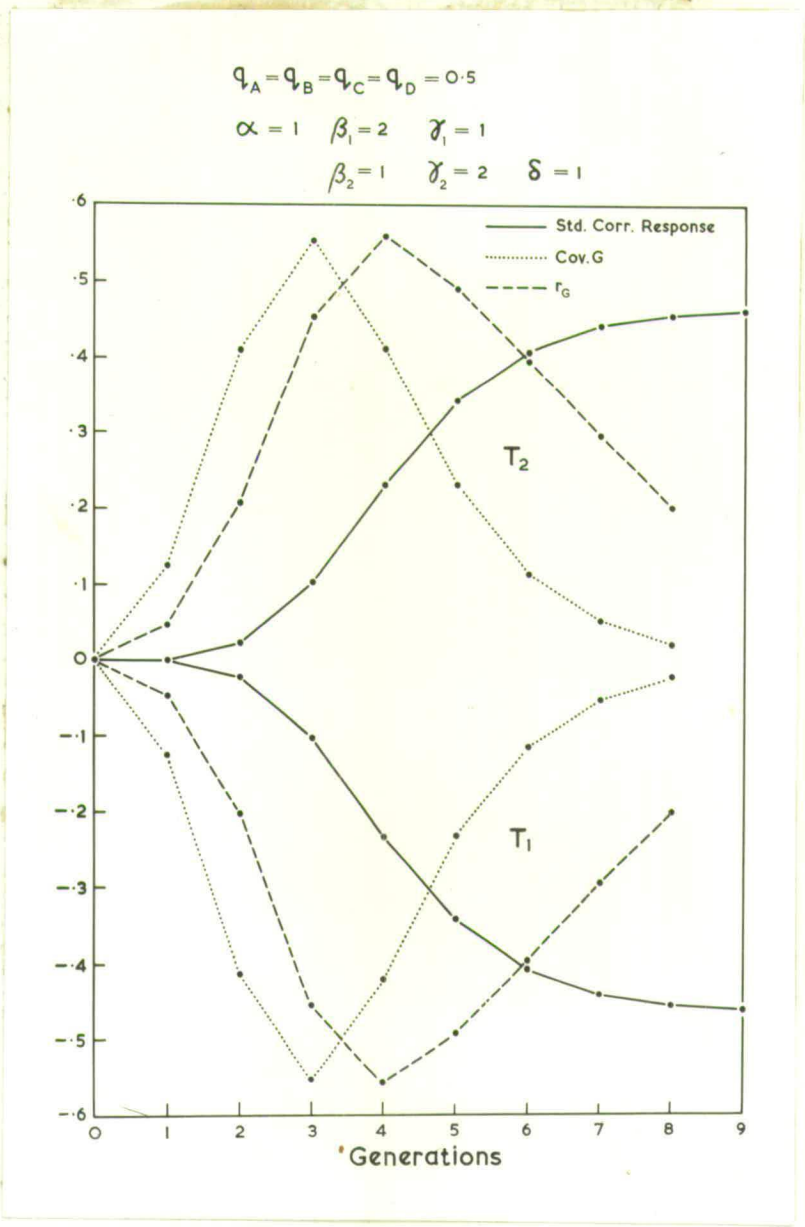


Figure 1 Standardized correlated responses, genetic covariances (cov G) and genetic correlations ($r_G$) for selection on trait 1 (T1) or trait 2 (T2). Model (vii)(a). All gene frequencies 0.5. Selection for trait 1 rapidly fixes a locus making a positive contribution to cov G, that for trait 2 fixes one making a negative contribution.

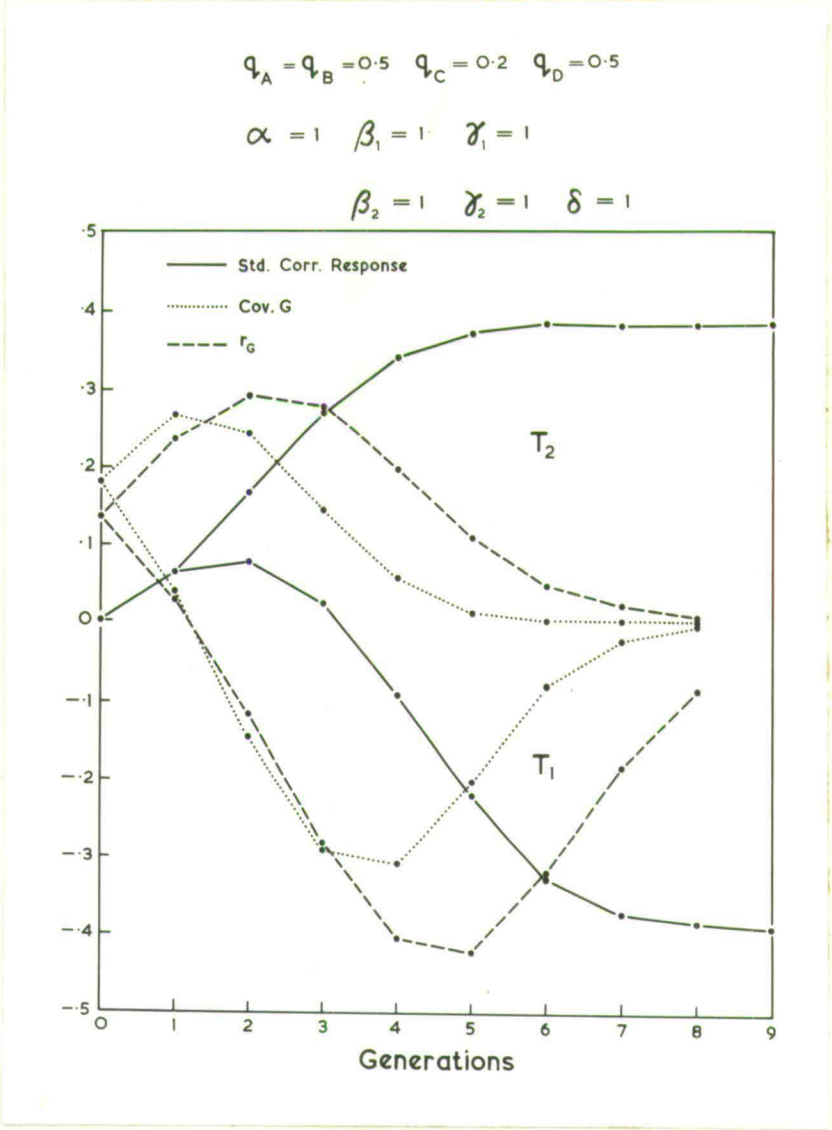


Figure 2 Standardized correlated responses, genetic covariances (cov G) and genetic correlations ($r_G$) for selection on trait 1 (T1) or trait 2 (T2). Model (i)(c). All gene effects equal. All frequencies 0.5 except that at a locus making a negative contribution to cov G. Perhaps the most frequent cause of asymmetry in practice?

(e) $q_B = q_C \neq 0.5$. Asymmetry in all models. Note that in (i), (ii) and (iii) there is no correlated response on selection for trait 1 though there is on selection for trait 2 (Fig. 3).

(f) and (g). Deviations from 0.5 in opposite directions in B and C loci. Asymmetry in all models.

The critical point is simply that asymmetry of the correlated responses occurs whenever the relative rate of response of the B and C loci is different when selection is for trait 1, than it is when selection is for trait 2. The combination of factors which can account for differing relative rates of change at these two types of loci when selection is for different traits are shown in equation (4). This equation is very powerful in the analysis of these correlated responses, and remarkably so considering that it is strictly valid for only a single generation of selection. The occurrence of symmetry was predictable from equation (4) in all models.

The table shows that quite remarkable degrees of asymmetry can be found in some of the models and differences in sign in the realized genetic correlation are frequently found, particularly in gene effect model (vii). Even with all gene frequencies at 0.5, the realized genetic correlation is about 0.25 for selection on trait 2 and -0.25 for trait 1.

Several computer runs were done with different heritabilities for the two characters ($h_1^2 \sim 2h_2^2$) and the results are given in Table 2. Gene effect model (i) still shows symmetry with all frequencies at 0.5. When the heritabilities were equal all four terms in (4) were zero, but now the linear terms are zero and the two square terms are equal but of opposite sign. But, with gene effect model (v), the change in heritability

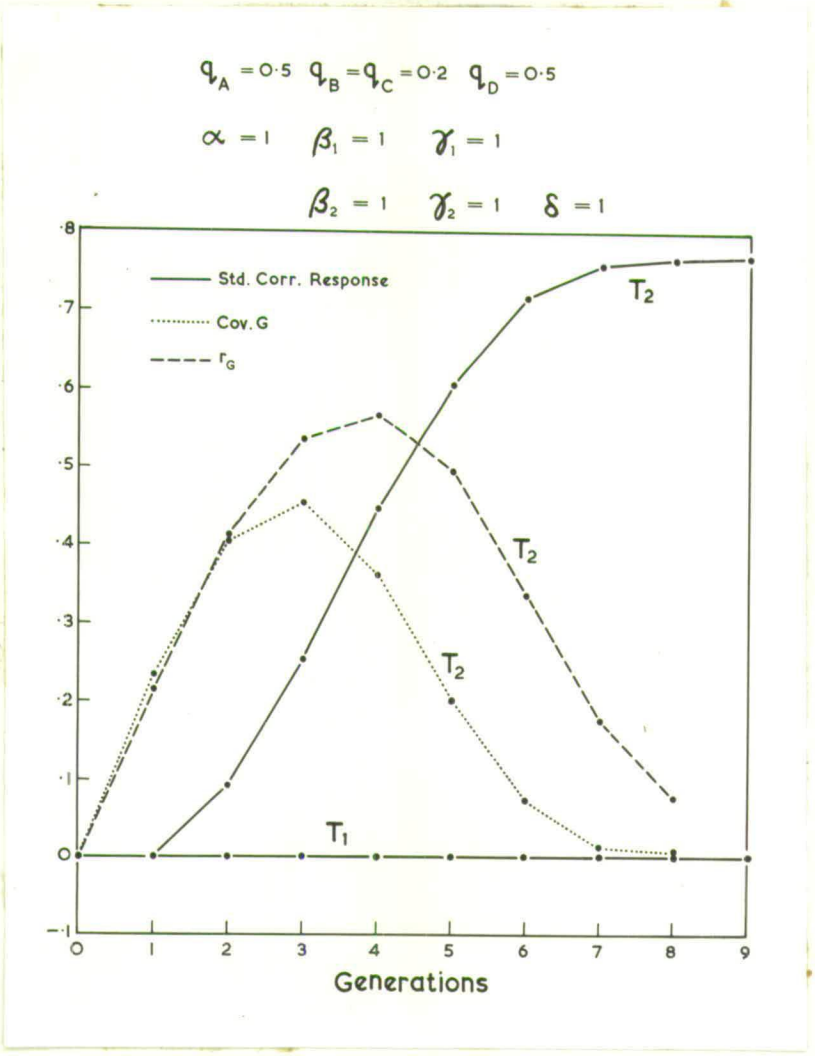


Figure 3. Standardized correlated responses, genetic covariances (cov G) and genetic correlations ($r_G$) for selection on trait 1 (T1) or trait 2 (T2). Model (i)(e). All gene effects equal. All frequencies 0.5 except at two loci, one with a positive and one with a negative contribution to cov G. Note the absence of any correlated response on selection for trait 1.

leads to asymmetry because the square terms in (4) are no longer zero. When both B and C loci have gene frequencies of 0.2, the change in heritability alters the existing asymmetry only a little.

In addition, one model with non-additive gene effects was studied. The model assumed complete dominance, equal gene effects at all loci and all gene frequencies at .25. This frequency was chosen because the absolute change in the mean in the first generation of selection would be the same whether the selection is up or down. This condition would be the most likely to yield symmetry. Even so, asymmetry after 9 generations was 0.21 standard deviations. Symmetry in the case of non-additive genetic effects could occur only if no negatively correlated loci were involved and the selective advantages of the positively correlated loci in the two traits were equal.

Selection in opposite directions for the same trait.

A similar method of analysis can be used to explain asymmetry of response in trait 2 when both up and down selection is practised for trait 1, and *vice versa*. With symmetry, the correlated responses in trait 2 should be of the same magnitude but of opposite sign, and asymmetry will be observed after the first generation only if there are parameter changes. From one generation of selection for trait 1, the difference between the genetic covariances after up selection (Cov $G_U$) and down selection (Cov $G_D$) turns out to be

$$\text{Cov } G_U - \text{Cov } G_D = \frac{2}{\sigma_1} \sum_{k=1}^{n} \lambda_k^2 \mu_k q_k (1-q_k) \left[ (1-2q_k)(\bar{I}_U + \bar{I}_D) - q_k(1-q_k) \frac{\lambda_k}{\sigma_1} (\bar{I}_U^2 - \bar{I}_D^2) \right] \quad (6$$

where the notation is the same as in equation (5), and $\bar{I}_U$ and $\bar{I}_D$ are the absolute values of the standardized selection differentials for up and down selection respectively. If $\bar{I}_U = \bar{I}_D = \bar{I}$, equation (6) reduces to

$$\text{Cov } G_U - \text{Cov } G_D = 4 \frac{\bar{I}}{\sigma_1} \sum_{k=1}^{n} \lambda_k^2 \mu_k q_k(1-q_k)(1-2q_k). \quad \ldots\ldots (7)$$

Equivalent formulae can be obtained when selection is practised on trait 2.

From formula (7), it can be seen that asymmetry is to be anticipated unless all gene frequencies are one-half, or in the situation where the changes in covariance due to genes with frequencies below 0.5 just balances that from genes at high frequencies. The result of Clayton et al. (1957) in which there was a positive correlated response to up selection, but none to down selection, could be explained by the presence of positively correlated genes at low initial frequency, with few or no negatively correlated genes.

It is quite possible for the correlated responses to be symmetrical on divergent selection for one trait but asymmetrical on selection for the other. Equal gene effects at the four types of loci and gene frequencies 0.5, 0.2, 0.2 and 0.5 would be an example of this.

Hazel (see Lerner, 1950, p.238) has pointed out that the eventual effect of simultaneous selection for two characters must be to reduce the genetic correlation by fixing first those loci contributing positively to the covariance. Some experimental support of this prediction has been presented by Friars, Bohren and McKean (1962) in poultry. Selection giving equal weight to one standard deviation in the two characters would give an expression for the change in genetic covariance after one generation of upward selection of

$$(\text{Cov } G) = \bar{I} \sum_{k=1}^{n} \lambda_k \mu_k q_k (1-q_k) \left[ (1-2q_k)\left(\frac{\lambda_k}{\sigma_1} + \frac{\mu_k}{\sigma_2}\right) - \frac{1}{2} q_k (1-q_k)\left(\frac{\lambda_k}{\sigma_1} + \frac{\mu_k}{\sigma_2}\right)^2 \right]$$

Obviously the loci with $\lambda$ and $\mu$ both of the same sign will contribute most to this change. But in early generations, the first term within the square

brackets may predominate and if such loci have low values of $q_k$, the genetic covariance may well increase for a while.

## DISCUSSION

Both from the algebraic treatment and from the computer results, it is clear that asymmetry of correlated response is likely to be found fairly frequently. The models are, of course, rather simplified and it should be asked what relevance these results have to real situations. The most obvious simplification is in the small number of loci in the models. The next degree of complexity would be to deal with n genes of each kind, but with the condition that the total additive genetic variance and the heritability of the two characters should remain the same. Then the scale of operations is altered by a factor $\sqrt{n}$, though the initial rate of response to selection will not be changed. The linear effects of the genes will be reduced by this factor, and the total advance under selection and the time scale of changes in the genetic parameters will be increased. If time is measured as a proportion of the total period of selection advance, the descriptions of asymmetry will become almost independent of the number of genes concerned. From the computer results it would seem that the greatest asymmetry (as measured by the difference between the genetic covariance in the two lines) will occur when the lines are about half-way to the final limit. The greater the number of genes concerned, the more likely it is that the terms linear in gene effects in equation (4) will be greater than those in which the effects are squared and the former will predominate in the early generations.

If time is measured in generations, then the greater the number

of genes concerned, the longer the time for the asymmetry to develop. The amount of asymmetry expected in the early generations of selection would be between $1/\sqrt{n}$ and $1/n$ times that in the original model with one locus of each type, depending on whether the linear or square terms in equation (4) contribute most to the asymmetry. The expression for the change in covariance on divergent selection for a single character has only linear terms in it and the effect would therefore be $1/\sqrt{n}$ times as large. The number of loci involved does not affect the eventual presence or absence of asymmetry. Unequal numbers of loci contributing positively or negatively to the covariance would have a similar effect on asymmetry as would unequal effects at the two loci in the model.

Nordskog and Festing (1962) have proposed a differential control of the genetic variance in the two characters (similar to model (iv)) and Siegel (1962) has proposed different heritabilities for the two characters as explanations of asymmetry of correlated response. From the results of this study, it appears that these causes will in some combinations lead to asymmetry, but that neither of these causes are, in themselves, either necessary or sufficient to produce asymmetry. The same is true of the gene frequencies at any one type of locus, the gene effects on the two traits at one locus, the ratio of the selection intensities in the two traits, and the ratio of the environmental variances in the two traits as used in the model studied. While there are many combinations of these factors which will lead to asymmetrical correlated responses (i.e., equation (4) $\neq 0$), only a few specific combinations of these factors will result in symmetrical correlated responses (equation (4) = 0). In our view, the most frequent combination of factors giving asymmetry

will be loci contributing negatively to the covariance and having gene frequencies other than 0.5.

Perhaps the most important consequence of these results is not directly concerned with the asymmetry itself. If asymmetry exists, any a priori prediction of correlated response must have been incorrect. It has been accepted in quantitative genetic theory that predictions of direct response have only short-term validity because of the necessary changes that selection would being about in the genetic variance. It appears from the results that the genetic covariance between two characters may be even more sensitive to changes in gene frequency brought about by selection, and presumably also to changes due to random sampling when the population size is small. The additive genetic variance of any character will be made up of contributions from the separate loci. These contributions will change as the gene frequencies are altered by selection or by random drift and they will not all change in the same way, depending on the gene frequencies at the loci concerned. But the genetic covariance (if the genetic correlation is not close to 1) will either be made up of a much smaller number of terms, if all loci contribute to the covariance with the same sign, or will be made up of positive and negative contributions from different loci. In either case the proportional change in the genetic covariance is likely to be greater than in the genetic variances themselves. It must therefore be expected that the static description of a population in terms of additive genetic variances and covariances will be valid in prediction over a much shorter period for correlated responses than it will be for direct responses.

If the patterns of correlated responses in any situation are

to be fully understood, it will be necessary to analyse the basic causes of the genetic correlations between characters. Our results point clearly to the need for the development of new experimental techniques for this purpose.

## SUMMARY

The pattern of changes of the genetic covariance between two characters on selection was examined in an effort to explain the asymmetry of correlated responses in two traits, or of the same trait in two environments, frequently observed in experimental results.

The algebraic conclusions were further examined by model selection experiments using a computer. The computer was programmed to calculate the change in gene frequency from generation to generation and to calculate from it the expected changes in genetic variances and covariance as selection proceeded. This procedure was carried out with several models of gene effects and gene frequencies.

Asymmetry of the genetic covariance, and consequently of the correlated responses, resulted when the relative change in gene frequency at the loci contributing positively and negatively to the covariance depended on the trait selected. The conditions necessary for the development of asymmetry were examined and the results suggest that any symmetry found in an experiment is perhaps more surprising than asymmetry. Probably the most frequent contribution to asymmetry in practice will be from loci contributing negatively to the covariance and having frequencies other than 0.5.

Accurate prediction of correlated response over many generations is therefore not possible without prior knowledge of the

composition of the genetic covariance, as well as its magnitude. The validity of existing theory for the prediction of correlated responses is likely to be much poorer than for the prediction of direct responses. Predictions would then have to be based on the genetic parameters estimated in each generation.

REFERENCES

Bell, A.E. & H.W. McNary (1963). Genetic correlation and asymmetry of the correlated response from selection for increased body weight of Tribolium in two environments. Proc. XIth Int. Cong. of Genetics, p.256.

Clayton, G.A., G.R. Knight, J.A. Morris & A. Robertson (1957). An experimental check on quantitative genetic theory. III. Correlated responses. J. Genetics, 55, 171-180.

Falconer, D.S. (1960a). Introduction to Quantitative Genetics. ix + 365 pp. Edinburgh: Oliver and Boyd.

Falconer, D.S. (1960b). Selection of mice for growth on high and low planes of nutrition. Genet. Res. Camb., 1, 91-113.

Friars, G.W., B.B. Bohren & H.E. McKean (1962). Time trends in estimates of genetic parameters in a population of chickens subjected to multiple objective selection. Poultry Sci., 41, 1773-1784.

Griffing, B. (1960). Theoretical consequences of truncation selection based on individual phenotype. Aust. J. Biol. Sci., 13, 307-343.

Lerner, I.M. (1950). Population Genetics and Animal Improvement: 342 pp. Cambridge University Press.

Nordskog, A.W. & M. Festing (1962). Selection and correlated responses in the fowl. Proc. XIIth World Poultry Congress section papers, pp.25-29.

Reeve, E.C.R. & F.W. Robertson (1953). Studies in quantitative inheritance. II. An analysis of a strain of Drosophila melanogaster selected for long wings. J. Genet., 51, 276-316.

Siegel, P.B. (1962). A double selection experiment for body weight and breast angle at eight weeks of age in chickens. Genetics, 47, 1313-1319.

Yamada, Yukio, & A.E. Bell (1963). Selection for 13 day larval growth in Tribolium under two nutritional levels. Proc. XIth Int. Cong. Genetics, p. 256.

# APPENDIX B

## SPECIALISED SIRE AND DAM LINES:

## IV SELECTION WITHIN LINES

Rom Moav* and W.G. Hill,

Institute of Animal Genetics, Edinburgh 9

* On leave from the Hebrew University, Jerusalem, Israel.

## Introduction

In the earlier papers of this series algebraic and graphical procedures were developed for determining the relative profitability of pure lines and crosses between them. It was found that many of the complex situations associated with this problem could be understood and solved more easily by graphical techniques.

In this paper we propose to extend the graphical method to the solution of problems of selection within lines and within specialised sire and dam lines. Smith (1964) studied the problem algebraically and concluded that selection in specialised sire and dam lines is at least as efficient as selection within a single line, and that the relative efficiency of the former increases if there is an unfavourable correlation between the two sets of traits under selection. We also propose to investigate the efficiency of selecting males and females on different indices within a single line.

In order to simplify the presentation of this paper, we will assume that (i) both males and females are selected with equal intensity and have the same generation interval; (ii) the same selection indices can be applied to both males and females, an assumption which cannot be realised in practice if one of the traits is reproductive performance, unless selection is based on relatives' performance; (iii) the two traits under selection are uncorrelated; (iv) the traits are genetically additive, and (v) the population parameters, other than the means, are the same for each line and do not change as a result of selection. The discussion here is restricted to

selection for productivity and reproductive performance in pigs and poultry, although the techniques employed can easily be extended to other animals.

Mathematical details have been excluded from the body of the paper and are given in the Appendix. Also in Appendix 3 are given some algebraic results for correlated traits, but for simplicity these are not discussed in the text.

Selection indices in a single line.

Smith (1964) and Moav and Moav (1966) expressed profit (P) from a unit of produce as a function of the productive efficiency (y) of the offspring and the reproductive performance (x) of their parents by the relation

$$P = C - G(y) - N(x) \qquad (1)$$

where C is a constant and G and N are functions. Assuming genetic additivity, then G and N may be replaced by constants and equation (1) becomes

$$P = C - Gy - N/x \qquad (2)$$

so that profit is directly proportional to productivity and inversely proportional to reproductivity.

For a given value of P, y can be expressed as a function of x, and therefore a graph of this function joins all genetic stocks with the same profitability. Such graphs are termed profit contours, and a collection of these contours a profit diagram. An example is given in Figure 1 for broilers, where profit in pence per pound live weight is given by

$$P = 10.6 - 0.1y - 320/x \qquad (3)$$

where y is market age in days and x is the total egg production per hen (Moav and Moav, 1966).

A profit diagram can be used to show how selection changes the profitability of a population. If a selection differential of i standard deviations is applied to x alone, then the genetic change in x, denoted $\Delta x$, is $ih_x^2\sigma_x$, where $h_x^2$ and $\sigma_x$ are the heritability and phenotypic standard deviation of x. If selection is practised on y alone with the same intensity, then $\Delta y = ih_y^2\sigma_y$. In Figure 1 it is assumed that $x = 108$, $\sigma_x = 20$ eggs, $h_x^2 = 0.1$ and $y = 70$, $\sigma_y = 4$ days, $h_y^2 = 0.25$ and the traits are uncorrelated. For this example, if a selection intensity of 3 standard deviations were applied to x, then the response would be $\Delta x = 6$ eggs, and the consequent change in profit ($\Delta P_x$) = 0.15 pence, calculated by linear interpolation between the profit contours of Figure 1. On the other hand, selection on y with the same intensity would give a response $\Delta y = -3$ days with a subsequent change in profit ($\Delta P_y$) of 0.3 pence.

Alternatively, animals may be selected on an index of the two traits (Hazel, 1943), in which phenotypic values are weighted by the formula

$$I = x + By \qquad (4)$$

For uncorrelated traits, the responses in the component traits from selection on I are

$$\Delta x = ih_x^2\sigma_x^2/\sigma_I, \quad \Delta y = iBh_y^2\sigma_y^2/\sigma_I \qquad (5)$$

where $\sigma_I = \sqrt{\sigma_x^2 + B^2\sigma_y^2}$ is the standard deviation of the index. For example, the change in the population mean as a result of selection on

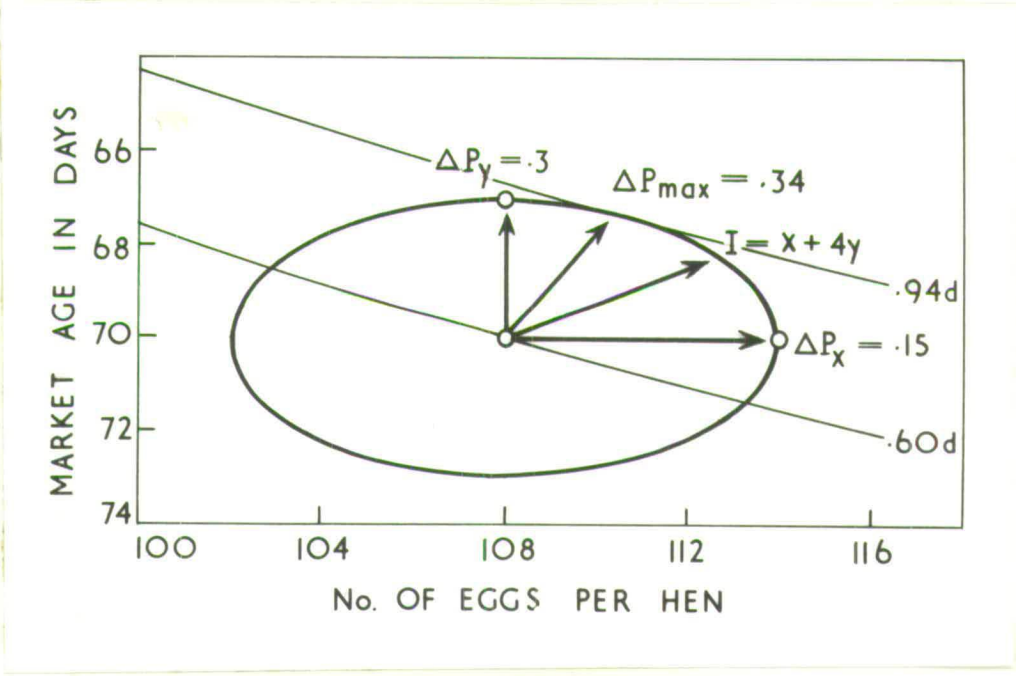


FIGURE 1 The profit diagram together with a response ellipse for broilers. Expected changes in profit are shown for selection based on market age alone ($\Delta P_y$), egg number alone ($\Delta P_x$) and on the optimum index ( $\Delta P_{max}$).

the index $I = x + 4y$, when $i = 3$, is given in Figure 1. It is evident that for a given set of parameters and intensity of selection, the index weight B determines the magnitude and direction of the responses. The locus of all points ( $\Delta x$, $\Delta y$) obtained by varying B, but with constant i, can be shown to be an ellipse (see Appendix 1). This ellipse will be termed the response ellipse, and is illustrated in Figure 1 for a selection intensity $i = 3$. It can be seen that the maximum profit from a given selection intensity is that of the highest profit contour which can be reached by the ellipse. At this point the ellipse is tangential to the profit contours. From a visual inspection of Figure 1 it can be seen that the maximum increase in profit for $i = 3$ is 0.34 pence. An algebraic method of using the response ellipse to determine the optimum index and change in profit is given in Appendix 1.

The use of an ellipse has some disadvantages however. In particular, the distance ( $\sqrt{(\Delta x)^2 + (\Delta y)^2}$) moved by the population mean on the profit diagram is dependent on the direction of selection, and some manipulation is required in order to compute index weights graphically from the ellipse. However, if the variables are transformed as follows

$$\Delta x^* = \frac{\Delta x}{h_x^2 \sigma_x}, \quad \Delta y^* = \frac{\Delta y}{h_y^2 \sigma_y} \qquad (6)$$

then the locus of the transformed variables ( $\Delta x^*$, $\Delta y^*$) is a circle, termed the response circle (see Appendix 2 for a proof). Since, for a selection intensity of i standard deviations applied to x, $\Delta x^* = i$, or, if applied to y, $\Delta y^* = i$, the transformed variables $\Delta x^*$ and $\Delta y^*$ can be regarded as standardised units of response of the traits to selection. In Figure 2, the response ellipse of Figure 1 has been

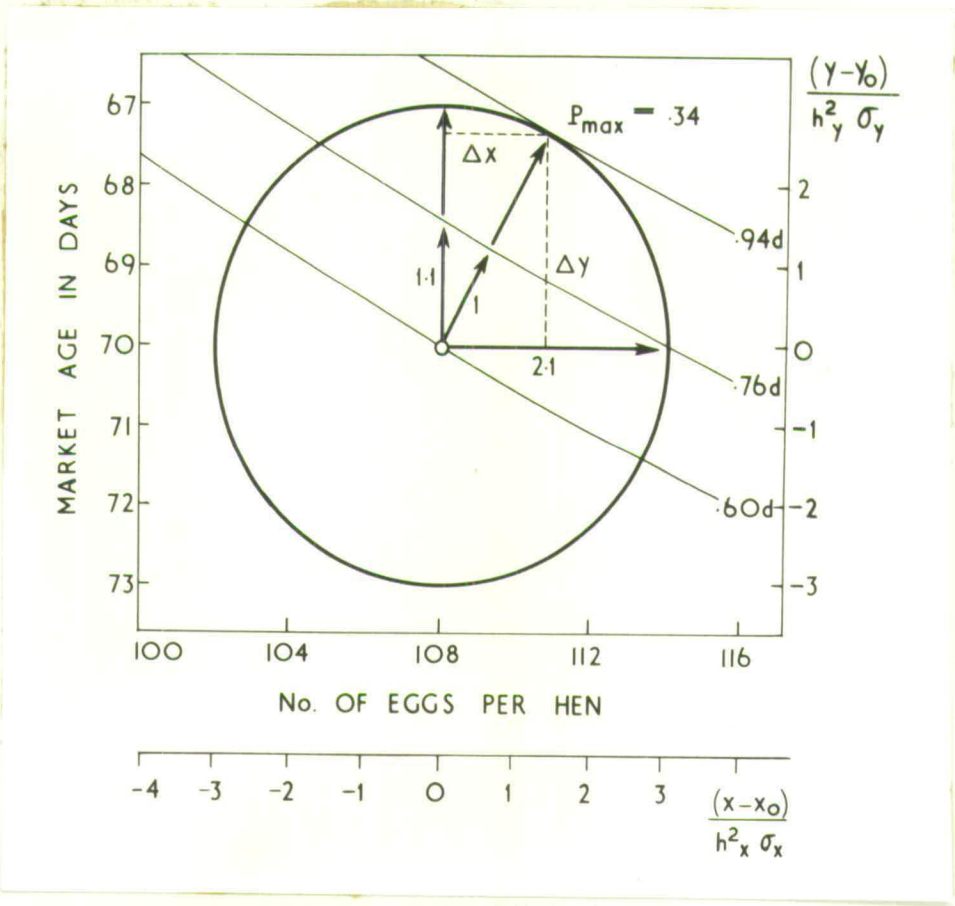


FIGURE 2 The example of Figure 1 transformed to a response circle. The lengths of the response vectors show the relative selection intensities required to make the same economic gains when selecting on market age alone, egg number alone and on the optimum index.

transformed into a response circle, which can be much more easily constructed than the ellipse. Again, the point of maximum profit is where the circle is tangential to the profit contours, and it also follows that the direction of response for maximum economic gain is perpendicular to the profit contours.

The response circle can be used to calculate the optimum index weighting B, from a profit diagram. The direction of the response for the optimum index is found by drawing a perpendicular to the contours, and for an arbitrary contour $\Delta x^*$ and $\Delta y^*$ measured. Then the index weight (see Appendix 2) is given by

$$B = \left(\frac{\Delta y^*}{\Delta x^*}\right)\left(\frac{\sigma_x}{\sigma_y}\right) \qquad (7)$$

In the example of Figure 2, $\Delta y^* / \Delta x^* = 1.9$ and $\sigma_x/\sigma_y = 5$, so that the optimum index is $I = x + 9.5y$. This result is the same as that obtainable by the algebraic methods of Hazel (1943).

On the standardised scale the length of the response vector (the line on the profit diagram joining the population means before and after selection) is a constant and equals i, the intensity of selection. Therefore, the efficiency of different selection indices can be compared by drawing the response vectors from the population mean to a convenient profit contour for different indices and measuring their lengths. For example, if the length of the vector of the most efficient index in Figure 2 is given a value of one, then it can be seen from the graph that 2.1 or 1.1 units of selection on x or y alone respectively would be needed to achieve the same increase in profit. Thus, in the present example, selection on y alone reduces efficiency by only 10%.

Similarly, it can be shown that changes in B over the range from 4 to 35 reduce efficiency not more than 5% below that of the optimum index, B = 9.5.

For the profit equation (2), the change in profit resulting from changes in reproductivity ($\partial P/\partial x$) and productivity ($\partial P/\partial y$) are given by

$$\frac{\partial P}{\partial x} = \frac{N}{x^2}, \qquad \frac{\partial P}{\partial y} = -G \qquad (8)$$

Since profit is linearly related to productivity, changes in y produce the same change in profit at all levels of y. On the other hand, since there is a non-linear relationship between profit and reproductivity changes in x yield changes in profit dependent upon the level of x. Thus for the profit equation (2), it can be seen from (8) that the higher the present level of reproductive performance, the greater the improvement necessary to produce the same increase in profit. These points are demonstrated in Figures 3A and 3B for pigs, for which the profit (P) in pence per pound live weight is

$$P = C - 3.4y - \frac{13000}{Wx} \qquad (9)$$

where y is the feed conversion ratio, W the market weight and x the number of pigs raised per sow per year (Moav, 1966). The constant C was not estimated, but it does not affect changes in profit or relative profitabilities. Figures 3A and 3B are drawn on a standardized scale ($h_x^2\sigma_x$, $h_y^2\sigma_y$), so that the lengths of the response vectors show the relative selection intensities necessary to achieve the same improvement in profit from four different levels of reproductivity, when selection is based on x alone, y alone, or on the optimum index. It can also be

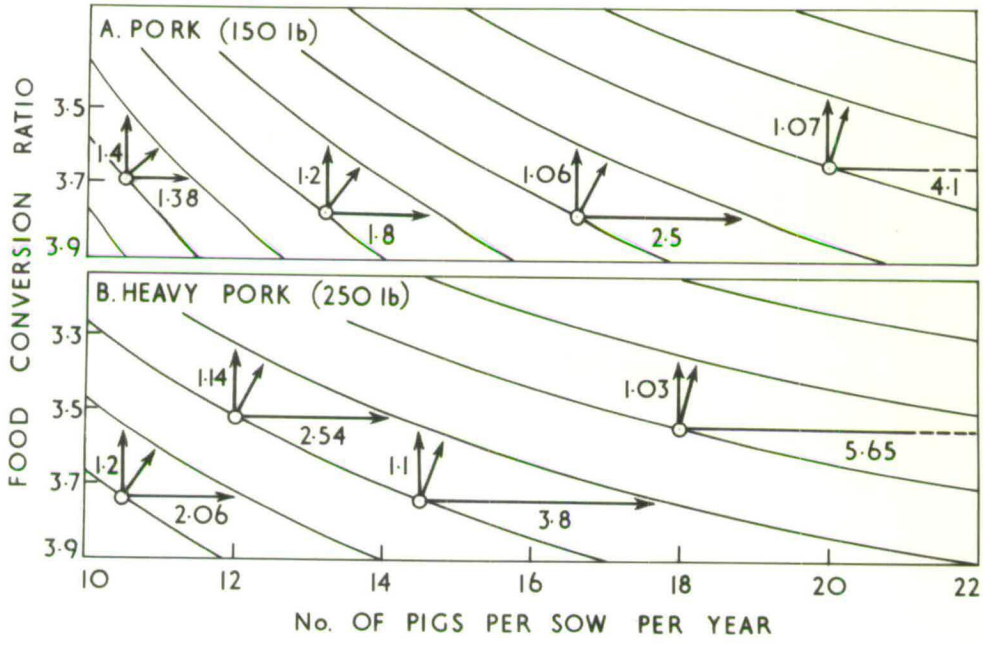


FIGURE 3 The effect of market age and level of reproductive performance on the efficiency of alternative selection schemes within single lines in pigs. The selection intensity needed to produce a given change in profit when selecting for food conversion ratio or reproductive performance alone is given relative to that necessary to make the same change in profit when selecting on the optimum index.

seen from Figure 3 that as the reproductive performance improves, less weight should be applied to it in a selection index, and it can be shown algebraically (see Appendix 4) that the optimum index for the profit equation (2) is given by

$$B = - \frac{x^2 G h_y^2}{N h_x^2} \qquad (10)$$

The differences between Figures 3A and 3B demonstrate the effect of market weight on the profit contours and selection responses. Since the maintenance cost of the sow is constant, as market weight increases the reproductive cost per pound of meat becomes increasingly smaller. Thus at higher levels of reproductive performance, selection on feed conversion efficiency (y) alone is almost as efficient as selection on the best index.

Selection in sire and dam lines

In most classes of livestock, variation in the reproductive performance of the male has a negligible effect on profitability and can be ignored (Smith, 1964; Moav, 1966). Assuming genetic additivity and independence of the component traits, then the profitability of the cross breds ($P_{SD}$) is a function of the arithmetic mean of the productivity of the two parents and the reproductive performance of the dam. Thus from equation (2)

$$P_{SD} = C - \frac{G}{2} (y_S + y_D) - \frac{N}{x_D} \qquad (11)$$

where the subscripts S and D refer to the sire and dam respectively. The sire and dam terms in (11) may be collected to show their contributions to costs:

$$P_{SD} = C - (\frac{G}{2} y_D - \frac{N}{x_D}) - \frac{G}{2} y_S$$

Since the sire line affects costs only through productivity, selection in the sire line should be based on that criterion alone. However, the contribution of the dam line to costs ($V_D$) is

$$V_D = \frac{G}{2} y_D - \frac{N}{x_D} \qquad (12)$$

so selection on the dam line should be on both traits and aimed at minimising $V_D$. The methods described earlier for single lines can be used, for profit contours can be drawn by expressing $y_D$ as a function of $x_D$ in (12), and a response circle constructed. The response vector of the most efficient dam line index is then the vector which is perpendicular to a dam contour.

The example of Figure 4 illustrates the construction of dam line indices, and shows how the profitability of crosses is affected by selection in their parental lines. This figure represents the hypothetical situation in which three lines of pigs are available. The line with the best productivity is chosen as the sire line, and denoted S; the other two lines are therefore alternative dam lines (D1 and D2). Figure 4 shows that, in this example, D1 and D2 have different performances as single lines, but they are equally profitable as dam lines. D1 and D2 fall on the same dam contour ($V_D$), and their crosses (SD1 and SD2) with the sire line also have the same profitability.

Also shown in Figure 4 are the responses to selection of the same intensity in the sire line, based on productivity, and in the dam lines, based on the appropriate optimal dam index. It is assumed that the genetic parameters, other than the means, are the same for each line.

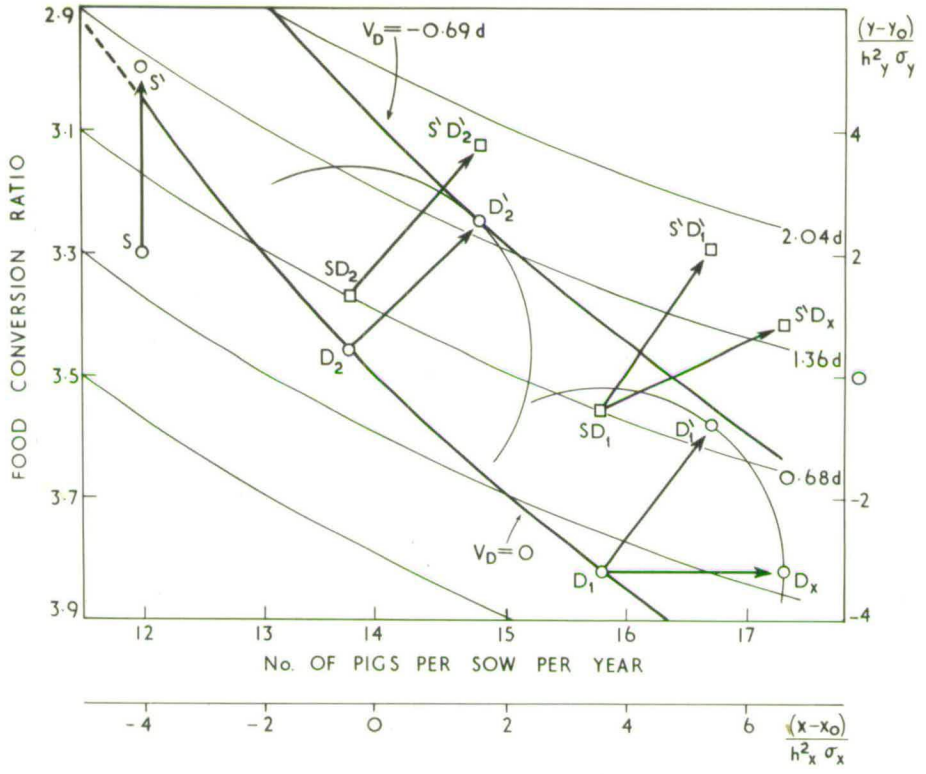


FIGURE 4 Selection within specialised sire (S) and dam ($D_1$, $D_2$) lines. The sire line is selected for food conversion ratio and the dam lines on the optimum index to reduce dam costs ($V_D$). The performances of the lines after selection are denoted S', $D_1'$, $D_2'$ and the profitability of their crosses $S'D_1'$, $S'D_2'$. Selection of $D_1$ for reproductive performance alone ($D_x$) and its cross ($S'D_x$) is also shown.

The new line means are denoted S', $D_1'$ and $D_2'$. Since the reproductive performance of $D_2$ was poorer than that of $D_1$, the improvement in profitability as a dam line resulting from selection has been rather greater in $D_2$ than in $D_1$. Similarly, the new cross $S'D_2'$ is more profitable than the cross $S'D_1'$. Whilst the differences in profitability of the dam lines after selection are small, the example illustrates that the ranking of lines can be affected by selection, even though the genetic parameters and selection intensities in such lines are the same.

Selection of males and females on specialised indices within a single line.

In this section we compare the efficiency of selecting males and females on different indices within a single line with the efficiency of selecting both sexes on the same index.

A graphical solution to this problem is shown in Figure 5, which has a standardized scale ($h_x^2\sigma_x$, $h_y^2\sigma_y$). The response circle from the original population, O, for one standard deviation of selection is drawn, and some specific response vectors marked. The vector I shows the response from selecting both sexes on the most efficient single line index. The vectors S and $D_x$ show the response from selection only on y and only on x respectively, and the vector $D_m$ the response when selection is based on the best dam line index. Thus the vector I is perpendicular to the profit contours (equation 2) and the vector $D_m$ perpendicular to the dam contours $V_D$ (equation 12). The profit of the progeny when selecting on males and females separately is at $SD_m$ and $SD_x$ for the dam selection vectors $D_m$ and $D_x$ respectively. It can be shown (see Appendix 5) that if the sires are selected only on y, and the dam index varied,

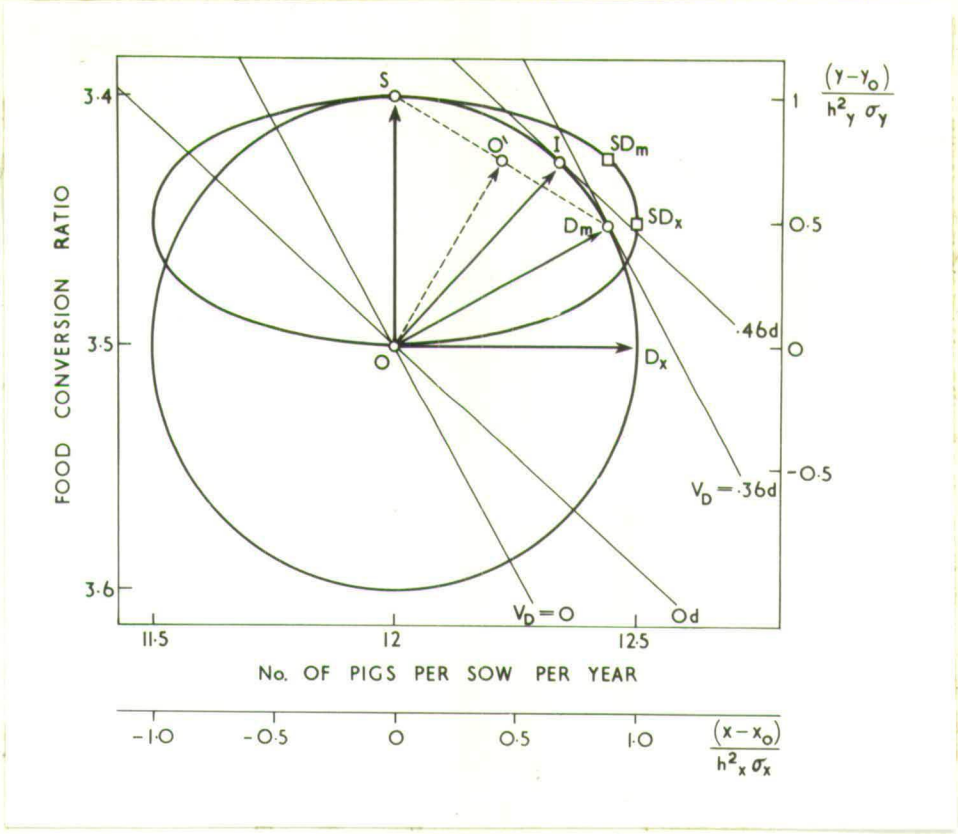


FIGURE 5 Selection of males and females on specialised indices within a single line. The population (O) is selected for food conversion ratio alone (S), reproductive performance alone ($D_x$), on the optimum single line index (I) perpendicular to the profit contours, and on the optimum dam line index ($D_m$) perpendicular to the dam line contours ($V_D$). The profitability of the progeny of sire S and dams $D_m$, $D_x$ are denoted $SD_m$, $SD_x$ respectively, and fall on the sire-dam response ellipse. The performance of the progeny S by $D_m$ is at O'.

then the profitability of the progeny (SD) falls on an ellipse termed the sire-dam response ellipse with co-ordinates $x = x_o$ and $y = y_o + \frac{\overline{S-O}}{2}$ where $x_o$ and $y_o$ are the co-ordinates of the base population and $\overline{S-O}$ the length of the sire response vector. We see from Figure 5 that the upper right quadrant of the sire-dam response ellipse lies outside the response circle. Thus we have a graphical proof of Smith's (1964) conclusion that selection on specialised indices is always at least as efficient as selecting both sexes on the same index.

We have considered the effects of different ways of selecting individuals on the profitability of their progeny. Let us now investigate the effects of these procedures on the profitability of subsequent generations. In order to do this we have to distinguish between the profitability of the progeny and their merit as parents for the next generation. Assuming additivity, their profitability has been shown to be a function of the mean productivity of both sets of parents and the reproductivity of only their dams. However, the population from which selection must now be practised has a performance which is the arithmetic mean of both parental traits. In the example of Figure 5, the point O represents the mean of a population formed from mating sires which were selected on y alone, and dams which were selected on the best dam index. O' is therefore the mid-point of the line S to $D_m$. Here S and $D_m$ represent the population means after selection on the vectors S and $D_m$ respectively; we shall use I similarly. As O' lies on a chord of the response circle, it must have a lower profitability than I, the progeny mean when parents of both sexes are selected on the same optimum index, which lies on the circumference of the response circle. Thus, since their parental

performance is poorer, the "grandchildren" will be poorer if their grandparents are selected on separate indices rather than on the same index, so that, although using a separate index maximises profit after one generation, it reduces profit in subsequent generations.

Discussion

There has been some indication in the profit diagrams that selection for reproductive rate leads to only small improvement in profitability for single lines (Figures 2 and 3) and even in specialised dam lines (Figure 4). There are two main causes: firstly, reproductive performance has a low heritability and, secondly, if reproductive rate is already fairly high, the economic returns from further improvement are small (equation 8). We shall now give a more detailed example to illustrate this point, and consider only feed costs which, of course, comprise the major portion of total costs in most livestock enterprises.

If the total food consumed per pound of live pig produced is F, then we can write

$$F = F_c + y + \frac{F_m}{Wx} \qquad (13)$$

where y is the food conversion efficiency of food eaten directly by the growing pig, $F_m$ is the food consumed by the dam per year for her own maintenance, W is the market weight and x the number of offspring reared per dam per year. $F_c$ is the food consumed by the dam in excess of maintenance during pregnancy and lactation, and depends on x. However, $F_c$ is relatively small and will be considered constant. Optimal selection indices can be constructed for (13) in the same manner as described earlier for equation (2).

This example is illustrated in Figure 6 for bacon pigs (W = 200 lbs)

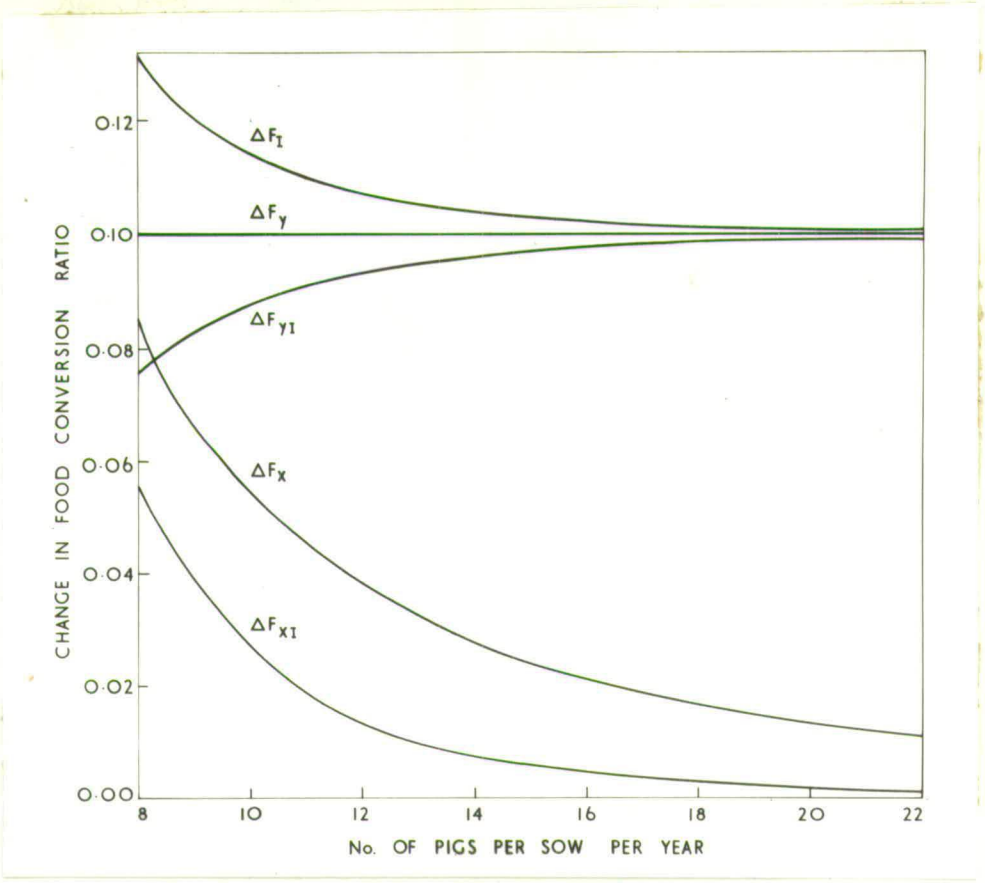


FIGURE 6 The effect of level of reproductive performance on the relative efficiencies of alternative selection schemes designed to minimise food costs in pigs, measured by the overall food conversion ratio which includes the maintenance cost of the dam. Selection is based on food conversion efficiency in the growing pig alone ($\Delta F_y$), on reproductive performance alone ($\Delta F_x$) and on the optimim index ($\Delta F_I$). The contributions ($\Delta F_{xI}$, $\Delta F_{yI}$) of the component traits to $\Delta F_I$ are also shown.

with the parameters $h_x^2 = 0.1$, $\sigma_x = 5$, $h_y^2 = 0.4$, $\sigma_y = 0.25$, and selection is in a single line. The graphs $\Delta F_x$ and $\Delta F_y$ show the changes in the overall feed conversion ratio F when selection is based on x or y alone. $\Delta F_I$ shows the change in F when selection is on the optimum index, and $\Delta F_{xI}$ and $\Delta F_{yI}$ are the corresponding changes in F resulting from changes in the component traits x and y, such that $\Delta F_I = \Delta F_{xI} + \Delta F_{yI}$. The example clearly shows that as the reproductive performance increases, little gain is made by selecting for x, that almost all the improvement $\Delta F_I$ comes from $\Delta F_{yI}$, and that the index is little more efficient than selecting on y alone.

In view of these conclusions, we may consider why so much emphasis is placed on selecting for reproductive performance in the commercial breeding of pigs and broilers. One possible reason is that throughout this series of papers we have assumed that the demand for the final produce is fixed, and that the number of parents is adjusted to meet this demand. This approach strictly holds only for national or regional evaluation, or for very large single enterprises. However, a small farmer may regard the number of sows, say, that he keeps as fixed, so that increasing their litter size not only reduces that part of the sow's maintenance cost to each piglet, but can also increase the total turnover without increasing capital expenditure. Since the small farmer is only supplying a very small part of the total demand, increases in his production would produce negligible effects on the price he receives for his product. Another reason why excessive selection pressure may be applied to reproduction is that, in a non-integrated system, the young animals may be bought from the multiplier by the grower at a price

dependent only on numbers and present weight or age, with little regard to their future performance. Thus as the dam costs contribute a much higher proportion of the total costs of the multiplier than of the grower, and since the multiplier is the breeder's direct customer, the latter is forced to exert extra selection pressure on reproduction.

Let us now return to a discussion of the relative merits of different procedures for selection from a single line. The three alternatives are illustrated in Figure 7:

A. Maintaining a single line and selecting all animals on the same index,

B. Maintaining a single line but selecting males and females on specialised indices,

C. Splitting the original line into separate sire and dam lines, each selected on a specialised index.

A selection intensity of one standard deviation has been applied to pigs of each sex in each of five generations, and it is assumed that no parameters, other than the population means, change. The example shows that method C is most efficient, followed by A and then B but the differences between the methods are small. Thus, as Smith (1964) has shown, the maintenance of separate sire and dam lines is theoretically the most efficient method of improvement. In addition, there are several other advantages to be gained by maintaining separate lines. Smith showed that separate sire and dam lines become more efficient when productive and reproductive traits are negatively correlated. Separate lines allow heterosis in component traits to be expressed. Finally,

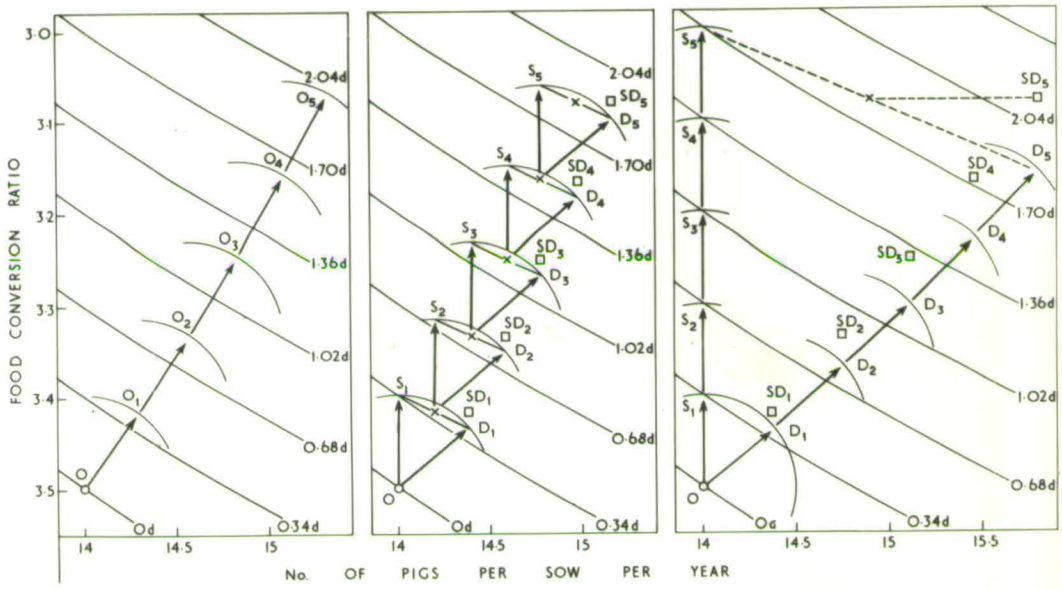


FIGURE 7 Long-term efficiency of alternative selection schemes. Left: Selection in a single line using the same index for each sex. Centre: Selection in a single line but using different indices for each sex. Right: Selection in specialized sire and dam lines. The fifth generation profitabilities are denoted $O_5$, $SD_5$ and $SD_5$ respectively.

with several lines, more genetic diversity is maintained as an insurance against changes in economic conditions or genetic parameters and against incorrect estimates of parameters at the beginning of the selection programme.

Summary

Selection on an index of two traits was represented graphically by means of a response ellipse or response circle. This procedure was used to find optimal index weights for uncorrelated traits, and to compare the efficiency of alternative indices.

Profit was expressed as a reciprocal function of reproductive performance, from which it was shown that the higher the reproductive performance, the smaller the weight that should be applied to it in a selection index. It was found that in pigs and broilers the average commercial standard of reproductive performance is sufficiently high that selection on production traits alone is almost as efficient as selecting on the optimum index.

Three alternative procedures for selecting from a single original line were compared graphically:

A. Maintaining the single line and selecting all animals on the same index.

B. Maintaining the single line but selecting males and females on separate indices.

C. Splitting the original line into separate sire and dam lines, each selected on a specialised index.

Method C was found to be the most efficient of the three; B was more efficient than A for one generation, but less efficient in subsequent generations.

Acknowledgements

The authors wish to thank Mr. R.B. Land and Dr. C. Smith for their critical reading of the manuscript and valuable suggestions.

One of us (W.G.H.) was in receipt of an Agricultural Research Council Postgraduate Studentship during the course of this work.

References

Hazel, L.N., 1943. The genetic basis for constructing selection indices. Genetics, 28: 476.

Moav, R., 1966. Specialised sire and dam lines. I: Economic evaluation of cross-breds. Anim. Prod.

Moav, R., and Moav, J., 1966. Profit in a broiler enterprise as a function of egg production of parent stocks and growth rate of their progeny. Br. Poul. Sci. (in press).

Smith, C., 1964. The use of specialised sire and dam lines in selection for meat production. Anim. Prod., 6: 337.

APPENDIX

1. The response ellipse

In the appendix we will find it convenient to define a more general notation than is used in the main part of the paper.

Let changes in profit be given by

$$\Delta P = a_x \Delta x + a_y \Delta y \qquad (1A)$$

where $a_x$ and $a_y$ are the partial regressions of profit on x and y respectively, and are usually called the economic weights. Let selection be practised on the basis of the index

$$I^o = b_x x + b_y y \qquad (2A)$$

For uncorrelated traits with heritabilities $h_x^2$, $h_y^2$, and phenotypic standard deviations $\sigma_x$, $\sigma_y$ then the variance of the index (2A) is

$$\sigma_{I^o}^2 = b_x^2 \sigma_x^2 + b_y^2 \sigma_y^2 \qquad (3A)$$

and the responses to selection with an intensity i standard deviations are

$$\Delta x = \frac{i b_x h_x^2 \sigma_x^2}{\sigma_{I^o}} \quad , \quad \Delta y = \frac{i b_y h_y^2 \sigma_y^2}{\sigma_{I^o}} \qquad (4A)$$

If equation (3A) is rewritten

$$\sigma_{I^o}^2 = \left( \frac{i b_x h_x^2 \sigma_x^2}{\sigma_{I^o}} \right)^2 \cdot \frac{\sigma_{I^o}^2}{(i h_x^2 \sigma_x)^2} + \left( \frac{i b_y h_y^2 \sigma_y^2}{\sigma_{I^o}} \right)^2 \cdot \frac{\sigma_{I^o}^2}{(i h_y^2 \sigma_y)^2} \qquad (5A)$$

and $\Delta x$ and $\Delta y$ are substituted from (4A) into (5A), and (5A) divided by $\sigma_{I^o}^2$ then

$$\frac{(\Delta x)^2}{(i h_x^2 \sigma_x)^2} + \frac{(\Delta y)^2}{(i h_y^2 \sigma_y)^2} = 1 \qquad (6A)$$

Equation (6A) describes an ellipse in the variables $\Delta x$ and $\Delta y$ and is termed the response ellipse. The ellipse has axes of length $ih_x^2\sigma_x$ and $ih_y^2\sigma_y$ which are parallel to the co-ordinate axes.

The point of maximum response and the weights of the optimum index are found by equating the tangents of the response ellipse and the profit contours. The tangent of a profit contour is $-a_x/a_y$, and the tangent of the ellipse (6A) is $\frac{-\Delta x}{\Delta y} \cdot \frac{h_y^4\sigma_y^2}{h_x^4\sigma_x^2}$. By equating tangents and rearranging, we obtain the optimum index

$$\frac{b_y}{b_x} = \frac{a_y h_y^2}{a_x h_x^2} \qquad (7A)$$

which is the same solution as can be obtained by Hazel's (1943) method.

If we define the index (equation (4) of the text) $I = x + By$, then for the optimum index

$$B = \frac{a_y h_y^2}{a_x h_x^2} \qquad (8A)$$

2. The response circle

If we define the transformations (equation (6) of the text)

$$\Delta x^* = \frac{\Delta x}{h_x^2\sigma_x}, \qquad \Delta y^* = \frac{\Delta y}{h_y^2\sigma_y}$$

then the ellipse (6A) reduces to

$$(\Delta x^*)^2 + (\Delta y^*)^2 = i^2 \qquad (9A)$$

Equation (9A) describes a circle of radius i, termed the response circle of the transformed variates.

If the economic weights of the transformed variables are denoted $a_x^*$ and $a_y^*$, then for

$$a_x^* \Delta x^* + a_y^* \Delta y^* = a_x \Delta x + a_y \Delta y$$

to hold for all $\Delta x$, $\Delta y$, we must define $a_x^*$ and $a_y^*$ by the inverse transformations

$$a_x^* = a_x h_x^2 \sigma_x, \quad a_y^* = a_y h_y^2 \sigma_y \qquad (10A)$$

Equating the tangents of the transformed profit contours and the response circle gives

$$\frac{\Delta y^*}{\Delta x^*} = \frac{a_y^*}{a_x^*} \qquad (11A)$$

If (8A) and (10A) are substituted into (11A) and the equation rearranged, we obtain formula (7) in the text

$$B = \frac{\Delta y^*}{\Delta x^*} \cdot \frac{\sigma_x}{\sigma_y}$$

## 3. Correlated traits

In this section we extend the theoretical results of Appendices 1 and 2 to two correlated traits. We shall use matrix algebra, and denote transposition by a prime (').

Let P be the phenotypic variance covariance matrix,

$$P = \begin{pmatrix} \sigma_x^2 & \text{Cov}\,(xy) \\ \text{Cov}\,(xy) & \sigma_y^2 \end{pmatrix},$$

let G be the genotypic variance covariance matrix, and let

$$\Delta = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}, \quad b = \begin{pmatrix} b_x \\ b_y \end{pmatrix}, \quad a = \begin{pmatrix} a_x \\ a_y \end{pmatrix}.$$

The selection responses on the index $I^o = b_x x + b_y y$ are

$$\Delta = iGb/\sigma_I o \qquad (12A)$$

where $\sigma_I^2 o = b' Pb$

Writing $b'Pb = b' GG^{-1} PG^{-1} Gb$

we can substitute (12A) to obtain the response ellipse

$$\frac{1}{i^2} \quad \Delta' G^{-1} PG^{-1} \Delta = 1 \qquad (13A)$$

By equating tangents, it can be shown that (13A) yields the same optimum index weights as does Hazel's (1943) method.

In order to transform (13A) to a circle, we define

$$M = \begin{pmatrix} \sigma_x \text{ Cos } \theta & \sigma_y \text{ Sin } \theta \\ \sigma_x \text{ Sin } \theta & \sigma_y \text{ Cos } \theta \end{pmatrix}$$

where Sin $2\theta$ = r, the phenotypic correlation. Since $M'M = P$, we can write (13A) in the form

$$\Delta' G^{-1} M'M G^{-1} \Delta = i^2 \qquad (14A)$$

so that if we define the transformed vector

$$\Delta^* = MG^{-1}\Delta$$

then $$(\Delta^*)' \Delta^* = i^2 \qquad (15A)$$

Equation (15A) is the formula of the response circle.

Similarly, we apply the inverse transformation to the profit contours

$$a^* = (M')^{-1} Ga,$$

and, after equating tangents and rearranging, we obtain the formula, analogous to equation (7) of the text, for computing index weights from

the response circle. This formula turns out to be

$$B = \frac{b_y}{b_x} = \frac{\sigma_x \left(1 + \sqrt{1-r^2}\right) \Delta y^* - r \Delta x^*}{\sigma_y \left(1 + \sqrt{1-r^2}\right) \Delta x^* - r \Delta y^*}$$

which, of course, reduces to (7) if $r = 0$.

4. Non linear profit contours

For the non linear profit contours of equation (2) we can substitute $a_x = N/x^2$, $a_y = -G$ (equation 8). Hence, from (8A) the optimum index is

$$B = \frac{b_y}{b_x} = \frac{-x^2 G h_y^2}{N h_x^2} \qquad (16A)$$

If, for brevity, we let

$$N^2 h_x^4 \sigma_x^2 = A_x, \quad G^2 h_y^4 \sigma_y^2 = A_y$$

then, from (3A), (4A) and (16A) the responses for the optimum index are

$$\Delta x = \frac{iA_x}{N\sqrt{A_x + x^4 A_y}}, \quad \Delta y = \frac{-iA_y x^2}{G\sqrt{A_x + x^4 A_y}}$$

The proportion of the change in profit due to changes in x is therefore, using (1A),

$$\frac{a_x \Delta x}{\Delta P} = \frac{A_x}{A_x + x^4 A_y}$$

and due to changes in y

$$\frac{a_y \Delta y}{\Delta P} = \frac{x^4 A_y}{A_x + x^4 A_y}$$

so that clearly, as the present reproductive performance (x) increases, a much greater proportion of the total economic response is contributed by improvement in productivity (y).

5. The sire-dam response ellipse

Let the selection intensities be $i_S$ and $i_D$ standard deviations in the sire and dam lines respectively. In the sire line selection is for productivity (y) only, so on the standardised scales (equation 6 of the text)

$$\Delta y_s^* = i_s .$$

In the dam selection is on both traits, so from (11A), we have

$$\Delta y_D^* = \sqrt{i_D^2 - (\Delta x_D^*)^2}$$

For the progeny (SD), therefore,

$$\Delta y_{SD}^* = (\Delta y_s^* + \Delta y_D^*)/2$$

$$\Delta x_{SD}^* = \Delta x_D^*$$

Combining the above equation, we obtain

$$\Delta y_{SD}^* = i_s/2 + (\sqrt{i_D^2 - (\Delta x_{SD}^*)^2})/2 \qquad (17A)$$

On rearrangement of (17A) we obtain the ellipse

$$\frac{(\Delta x_{SD}^*)^2}{i_D^2} + \frac{(\Delta y_{SD}^* - i_s/2)^2}{(i_D/2)^2} = 1 \qquad (18A)$$

Equation (18A) is termed the sire-dam response ellipse. It has a centre at $(0 , i_s/2)$ relative to the base population, and axes of length $i_D$ and $i_D/2$, for x and y respectively, parallel to the co-ordinate axes. The example in the text has $i_s = i_D$.