ACCOMMODATION OF LINKAGE IN MASS SELECTION THEORY

By B. GRIFFING*

[*Manuscript received May* 19, 1960]

Summary

The purpose of this paper is to develop mass selection theory which will accommodate not only linkage but will provide for different recombination frequencies in the two sexes.

The theoretical aspects of the linkage problem are developed in three stages:

(1) The mass selection theory for two loci is extended to accommodate different recombination values for the two sexes.

(2) A method is developed by which the generalized two-locus model may be used to cope with genetic situations which are considerably more complex. This method requires the estimation of the recombination value averaged over all possible pairs of loci.

(3) The expectations of the half-sib and full-sib covariances for a random-mating population are generalized to permit different recombination values for the two sexes. This allows unbiased estimates of genotypic variance components to be obtained.

Finally, application of the more general mass selection theory to the problem of detecting the influence of natural selection in modifying the effectiveness of artificial selection, is discussed.

I. INTRODUCTION

This paper is the second of a series in which the main objective is to generalize mass selection theory to include epistasis and linkage. Except for an abbreviated excursion by Kimura (1958), there has been no attempt to extend the mathematical theory of selection to include these phenomena. That is, no one has seriously attempted to explore selection theory using Kempthorne's (1954) generalized gene model which permits an exact treatment of epistasis. Likewise, no one has seriously attempted to solve the complex problem of linkage. Therefore, the past treatment of selection theory is inadequate, since, obviously, both epistasis and linkage are very real phenomena which should not be ignored.

In the first paper of this series (Griffing 1960), a hierarchical classification of hereditary units was considered. These units were the (1) gene, (2) gamete, and (3) individual. By using successively higher levels of hereditary units, successively higher levels of generalization of the selection theory were obtained. For example, when the individual was used as a unit of inheritance, the problem of linkage was avoided and certain very general statements were possible. However, by far the most informative approach was that in which the gamete was considered as the basic unit of inheritance, and the interpretation of the analysis was based on the gene.

* Division of Plant Industry, C.S.I.R.O., Canberra.

This gamete–gene approach is applied in the present paper. However, with this approach the linkage problem cannot be avoided and, in fact, since many experimental selection studies are conducted with *Drosophila*, it is necessary to handle not only the general concept of linkage, but it is also necessary to allow for the existence of different recombination values in the two sexes.

One approach to the linkage problem is to attempt to completely describe the genetic complexities (including all linkage parameters) in successively more complicated genetic systems (i.e. systems involving 2, 3, 4, . . . , n loci). However, the algebra quickly becomes intractable.

In this study the linkage problem is attacked, first, by deriving the theoretical consequences of selection when a two-locus model is used which is generalized to accommodate different recombination values for the two sexes. This two-locus model is then adapted to approximately describe a complex situation by simply replacing the specific recombination value for the two loci by the recombination value *averaged over all possible pairs of loci*. It turns out that a sufficiently accurate estimate of this average recombination value can be obtained by a simple expression which is a function of only the recombination index. Finally, the problem of estimating certain variance components from covariances is solved by extending the covariance formulae to accommodate different recombination values in the two sexes.

It is immediately obvious from the general extension of the mass selection theory that certain epistatic effects cause results which mimic those due to the effects of natural selection. Therefore, a method is outlined with which it is possible to detect the influence of natural selection in modifying the responses to artificial selection even when epistasis is present.

II. Consequences of Truncation Selection for the Completely Generalized Two-locus Model

In the first paper of the series (Griffing 1960), the consequences of truncation selection based on the individual phenotype were examined in detail for the two-locus model which was completely general except for the fact that the recombination value was assumed to be the same for the two sexes. With regard to this assumption, the following conclusions were drawn:

> "This simplification often does not exist. For example, an extreme case occurs in *Drosophila* where crossing over does not occur in the male. Thus, a somewhat more complicated analysis is required to accommodate different recombination values in the two sexes. However, such a complication does not change the general picture; it merely alters the speed of the response to selection and response to relaxation following selection."

In this section, then, the objective is to set out the analyses for the two-linked-locus case in which an arbitrary recombination frequency exists for each sex.

(*a*) *Definitions*

It is assumed that the selection programme commences with a random-mating population which is in equilibrium. This population is designated as Π_0. The following notation is used (Kempthorne 1957):

Let

$$\sum_i p_i^1(A_i^1) = \text{array of alleles at locus (1)},$$

$$\sum_{ij} p_i^1 p_j^1(A_i^1 A_j^1) = \text{genotypic array at locus (1)},$$

$$\sum_k p_k^2(A_k^2) = \text{array of alleles at locus (2)},$$

$$\sum_{kl} p_k^2 p_l^2(A_k^2 A_l^2) = \text{genotypic array at locus (2)},$$

y_f = recombination frequency between the two loci as exhibited by the females,

and

y_m = recombination frequency between the two loci as exhibited by the males.

The initial equilibrium population may be generated as the product of the two genotypic arrays, i.e.

$$\Pi_0 = [\sum_{ij} p_i^1 p_j^1(A_i^1 A_j^1)][\sum_{kl} p_k^2 p_l^2(A_k^2 A_l^2)],$$

$$= \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2(A_i^1 A_j^1 A_k^2 A_l^2).$$

Consider, now, the gametic arrays for each of the two sexes in Π_0. The female gametic array will be obtained first.

A female of the genotype $(A_i^1 A_k^2)(A_j^1 A_l^2)$ produces the following gametic array:

$$\{[(1-y_f)/2](A_i^1 A_k^2 + A_j^1 A_l^2) + (y_f/2)(A_i^1 A_l^2 + A_j^1 A_k^2)\}.$$

Hence, the total gametic array for the females is

$$\sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2\{[(1-y_f)/2](A_i^1 A_k^2 + A_j^1 A_l^2) + (y_f/2)(A_i^1 A_l^2 + A_j^1 A_k^2)\}.$$

This may be recast, using a summation device introduced by Kempthorne (1957), as follows:

$$\sum_{ijkl} \{p_i^1 p_j^1 p_k^2 p_l^2[(1-y_f)/2] + p_j^1 p_i^1 p_l^2 p_k^2[(1-y_f)/2] + p_i^1 p_j^1 p_l^2 p_k^2(y_f/2) + p_j^1 p_i^1 p_k^2 p_l^2(y_f/2)\}(A_i^1 A_k^2)$$

$$= \sum_{ik} ({}_f f_{ik}^0)(A_i^1 A_k^2),$$

where

${}_f f_{ik}^0$ = the relative frequency of the gamete $(A_i^1 A_k^2)$ produced by the females in Π_0

$$= p_i^1 p_k^2 \sum_{jl} \{[(1-y_f)/2](2p_j^1 p_l^2) + (y_f/2)(2p_j^1 p_l^2)\}$$

$$= p_i^1 p_k^2.$$

Likewise, the frequency of the gamete $(A_j^1 A_l^2)$ produced by the males in Π_0 is

$${}_m f_{jl}^0 = p_j^1 p_l^2.$$

Thus, when the population is in equilibrium the frequency of a particular gamete produced by either males or females is simply the product of the appropriate component gene frequencies.

It is now possible to give another representation of the random-mating population in equilibrium in terms of the gamete as a unit of inheritance. This representation is that obtained by multiplying the gametic arrays from the two sexes as follows:

$$[\sum_{ik}(_{f}f^{0}_{ik})(A^{1}_{i}A^{2}_{k})][\sum_{jl}(_{m}f^{0}_{jl})(A^{1}_{j}A^{2}_{l})] = \sum_{ijkl}(_{f}f^{0}_{ik})(_{m}f^{0}_{jl})(A^{1}_{i}A^{2}_{k})(A^{1}_{j}A^{2}_{l})]$$

$$= \sum_{ijkl} p^{1}_{i}p^{1}_{j}p^{2}_{k}p^{2}_{l}(A^{1}_{i}A^{1}_{j}A^{2}_{k}A^{2}_{l}),$$

as before.

The genotypic value of $(A^{1}_{i}A^{2}_{k})(A^{1}_{j}A^{2}_{l})$ in Π_0 is denoted as $d_{ik.jl}$, such that

$$\sum_{ijkl} p^{1}_{i}p^{1}_{j}p^{2}_{k}p^{2}_{l}\,d_{ik.jl} = 0.$$

This genotypic value is characterized by the following model (Kempthorne 1957):

$$d_{ik.jl} = \alpha^{1}_{i}+\alpha^{1}_{j}+\alpha^{2}_{k}+\alpha^{2}_{l}+\delta^{1}_{ij}+\delta^{2}_{kl}+(\alpha\alpha)_{ik}+(\alpha\alpha)_{il}+(\alpha\alpha)_{jk}+(\alpha\alpha)_{jl}$$
$$+(\alpha\delta)_{ikl}+(\alpha\delta)_{jkl}+(\delta\alpha)_{ijk}+(\delta\alpha)_{ijl}+(\delta\delta)_{ijkl},$$

where

α^{a}_{u} = additive genetic effect of the A^{a}_{u} allele,

δ^{a}_{uv} = dominance effect associated with the $A^{a}_{u}A^{a}_{v}$ genotype,

$(\alpha\alpha)_{ik}$ = additive × additive epistatic effect associated with genes A^{1}_{i} and A^{2}_{k},

$(\alpha\delta)_{ikl}$ = additive × dominance epistatic effect associated with the gene A^{1}_{i} and the genotype $A^{2}_{k}A^{2}_{l}$, and

$(\delta\delta)_{ijkl}$ = dominance × dominance epistatic effect associated with the genotypes $A^{1}_{i}A^{1}_{j}$ and $A^{2}_{k}A^{2}_{l}$.

The total genotypic variance may be partitioned as

$$\sigma^{2}_{G} = \sigma^{2}_{A}+\sigma^{2}_{D}+\sigma^{2}_{AA}+\sigma^{2}_{AD}+\sigma^{2}_{DD},$$

where

σ^{2}_{G} = total genotypic variance generated by the two loci,

σ^{2}_{A} = additive genetic variance,

σ^{2}_{D} = dominance variance,

σ^{2}_{AA} = additive × additive variance,

σ^{2}_{AD} = additive × dominance variance,

and

σ^{2}_{DD} = dominance × dominance variance.

In defining the selection value $w_{ik:jl}$ for the genotype $(A_i^1A_k^2)(A_j^1A_l^2)$ it is assumed that the genotypic variability of the characteristic which is being studied is controlled by genes, each of small effect, at many loci, and that the phenotypic varaibility is normally distributed with mean zero and variance σ^2. Following Kimura (1958), the selection value $w_{ik.jl}$ is defined to be proportional to the probability that an individual of the genotype $(A_i^1A_k^2)(A_j^1A_l^2)$ survives selection. Hence

$$w_{ik.jl} = 1+(i/\sigma^2)d_{ik.jl},$$

where i is the selection differential. Details of the argument are presented in the earlier paper (Griffing 1960).

(b) Consequences of n *Generations of Continuous Selection*

The objective in this section is to describe the change in parameters which occurs with an arbitrary number of continuous cycles of selection. The procedure will be to outline briefly the method of obtaining the population mean which results from one generation of selection and then to consider the consequences of n consecutive cycles of selection.

The selection programme starts with a random-mating population in equilibrium, as described in the previous section. This population is designated as Π_0, and the populations resulting from successive cycles of selection are designated as Π_i $(i = 1, \ldots, n)$.

The first cycle starts with

$$\Pi_0 = \sum_{ijkl} ({}_ff_{ik}^0)({}_mf_{jl}^0)(A_i^1A_k^2)(A_j^1A_l^2).$$

The frequency (male or female) of the genotype $(A_i^1A_k^2)(A_j^1A_l^2)$ following selection is

$$({}_ff_{ik}^0)({}_mf_{jl}^0)w_{ik.jl}^0.$$

The total frequency of the selected individuals (male or female) is

$$\sum_{ijkl} ({}_ff_{ik}^0)({}_mf_{jl}^0)[1+(i/\sigma^2)d_{ik.jl}^0] = 1.$$

The first step in obtaining the mean of Π_1 is to determine the frequency of a given gamete for each sex in the selected population. These give rise to gametic arrays for the selected males and females. The progeny mean, μ_1, is then obtained by multiplying these gametic arrays and substituting the genotypic value for the genotypes. In the following, the frequency of the female gamete $(A_i^1A_k^2)$ will be obtained first. The objective is to determine the gametic frequency as a function of the parameters of Π_0.

The female genotype $(A_i^1A_k^2)(A_j^1A_l^2)$ produces the following gametic array:

$$\{[(1-y_f)/2](A_i^1A_k^2+A_j^1A_l^2)+(y_f/2)(A_i^1A_l^2+A_j^1A_k^2)\}.$$

The total gametic array for the selected females is then,

$$\sum_{ijkl} ({}_f f^0_{ik})({}_m f^0_{jl}) w^0_{ik.jl}\{[(1-y_f)/2](A^1_i A^2_k + A^1_j A^2_l) + (y_f/2)(A^1_i A^2_l + A^1_j A^2_k)\} = \sum_{ik} ({}_f f^1_{ik})(A^1_i A^2_k),$$

where

$$\begin{aligned} {}_f f^1_{ik} &\cong [(1-y_f)/2][({}_f f^0_{ik})\sum_{jl}({}_m f^0_{jl}) w^0_{ik.jl} + ({}_m f^0_{ik})\sum_{jl}({}_f f^0_{jl}) w^0_{ik.jl}] \\ &\quad + (y_f/2)[\sum_{jl}({}_f f^0_{il})({}_m f^0_{jk}) w^0_{il.jk} + \sum_{jl}({}_f f^0_{jk})({}_m f^0_{il}) w^0_{jk.il}] \\ &= [(1-y_f)/2][({}_f f^0_{ik}) + ({}_m f^0_{ik})] + (i/\sigma^2) p^1_i p^2_k [\alpha^1_i + \alpha^2_k + (\alpha\alpha)_{ik}] + y_f p^1_i p^2_k \\ &= p^1_i p^2_k + (i/\sigma^2) p^1_i p^2_k [\alpha^1_i + \alpha^2_k + (\alpha\alpha)_{ik}]. \end{aligned}$$

Likewise, the frequency of the gamete $(A^1_j A^2_l)$ produced by the males is

$$\begin{aligned} {}_m f^1_{jl} &= [(1-y_m)/2][({}_f f^0_{jl}) + ({}_m f^0_{jl})] + (i/\sigma^2) p^1_j p^2_l [\alpha^1_j + \alpha^2_l + (\alpha\alpha)_{jl}] + (y_m) p^1_j p^2_l \\ &= p^1_j p^2_l + (i/\sigma^2) p^1_j p^2_l [\alpha^1_j + \alpha^2_l + (\alpha\alpha)_{jl}]. \end{aligned}$$

The structure of the population Π_1 may now be written as

$$\Pi_1 = \sum_{ijkl} ({}_f f^1_{ik})({}_m f^1_{jl})(A^1_i A^2_k)(A^1_j A^2_l),$$

which has the mean

$$\mu_1 = \sum_{ijkl} ({}_f f^1_{ik})({}_m f^1_{jl}) d^0_{ik.jl}.$$

This mean is approximately equal to

$$\sum_{ijkl}\{p^1_i p^1_j p^2_k p^2_l + (i/\sigma) p^1_i p^1_j p^2_k p^2_l [\alpha^1_i + \alpha^1_j + \alpha^2_k + \alpha^2_l + (\alpha\alpha)_{ik} + (\alpha\alpha)_{jl}]\} d^0_{ik.jl} = (i/\sigma)[\sigma^2_A + \tfrac{1}{2}\sigma^2_{AA}].$$

The approximation results from, first, making the transformation $i = i/\sigma$ and, then, assuming that the term

$$\left(\frac{\text{gene effect}}{\text{total phenotypic standard deviation}}\right)$$

is small, so that the square or product of two such quantities can be neglected. Assumptions of this sort are made throughout this analysis.

The consequences of n generations of continuous selection may be outlined briefly as follows:

The population resulting from $(n-1)$ consecutive cycles of selection has the following structure

$$\Pi_{n-1} = \sum_{ijkl} ({}_f f^{n-1}_{ik})({}_m f^{n-1}_{jl})(A^1_i A^2_k)(A^1_j A^2_l),$$

The gametic array produced by the selected females is then

$$\sum_{ik}({}_f f_{ik}^n)(A_i^1 A_k^2),$$

where

$$\begin{aligned}
{}_f f_{ik}^n &\cong [(1-y_f)/2][({}_f f_{ik}^{n-1})\sum_{jl}({}_m f_{jl}^{n-1}) w_{ik.jl}^0 + ({}_m f_{ik}^{n-1})\sum_{jl}({}_f f_{jl}^{n-1}) w_{ik.jl}^0] \\
&\quad + (y_f/2)[\sum_{jl}({}_f f_{ik}^{n-1})({}_m f_{jk}^{n-1}) w_{ik.jl}^0 + \sum_{jl}({}_f f_{jk}^{n-1})({}_m f_{il}^{n-1}) w_{jk.il}^0] \\
&= [(1-y_f)/2][({}_f f_{ik}^{n-1}) + ({}_m f_{ik}^{n-1})] + (i/\sigma^2) p_i^1 p_k^2 [\alpha_i^1 + \alpha_k^2 + (\alpha\alpha)_{ik}] \\
&\quad + (y_f) p_i^1 p_k^2 + (n-1)(y_f)(i/\sigma^2) p_i^1 p_k^2 (\alpha_i^1 + \alpha_k^2).
\end{aligned}$$

On converting to parameters of Π_0 only, it appears that the above frequency can be put in the form

$$\begin{aligned}
{}_f f_{ik}^n &= [(1-y_f) b^{n-1} + y_f + a(1-y_f) \sum_{r=1}^{n-1} b^{r-1}] p_i^1 p_k^2 \\
&\quad + [1 + (1-y_f) \sum_{r=1}^{n-1} b^{r-1}](i/\sigma^2) p_i^1 p_k^2 [\alpha_i^1 + \alpha_k^2 + (\alpha\alpha)_{ik}] \\
&\quad + [(n-1)(y_f) + \sum_{t=1}^{n-1} (\sum_{r=1}^{t} a(1-y_f) b^{r-1})](i/\sigma^2) p_i^1 p_k^2 (\alpha_i^1 + \alpha_k^2) \\
&= p_i^1 p_k^2 + n(i/\sigma^2) p_i^1 p_k^2 (\alpha_i^1 + \alpha_k^2) + \{1 + (1-y_f)[(1-b^{n-1})/(1-b)]\}(i/\sigma^2) p_i^1 p_k^2 (\alpha\alpha)_{ik},
\end{aligned}$$

where

$$a = \tfrac{1}{2}(y_f + y_m),$$

and

$$b = 1 - a.$$

In a similar manner the frequency of the gamete $(A_j^1 A_l^2)$ produced by the selected males is

$${}_m f_{jl} = p_j^1 p_l^2 + n(i/\sigma^2) p_j^1 p_l^2 (\alpha_j^1 + \alpha_l^2) + \{1 + (1-y_m)[(1-b^{n-1})/(1-b)]\}(i/\sigma^2) p_j^1 p_l^2 (\alpha\alpha)_{jl}.$$

Therefore, the mean of the population having n consecutive generations of selection can be determined as follows:

$$\begin{aligned}
\mu_n &\cong \sum_{ijkl}({}_f f_{ik}^n)({}_m f_{jl}^n) d_{ik.jl}^0 \\
&= (i/\sigma^2) n\sigma_A^2 + [(1-b^n)/(1-b)](i/\sigma^2)\tfrac{1}{2}\sigma_{AA}^2 \\
&= (i/\sigma^2) n\sigma_A^2 + (\sum_{r=1}^{n} b^{r-1})(i/\sigma^2)\tfrac{1}{2}\sigma_{AA}^2 \\
&= (i/\sigma^2)(\sigma_A^2 + \tfrac{1}{2}\sigma_{AA}^2) + (i/\sigma^2)[\sigma_A^2 + (b)\tfrac{1}{2}\sigma_{AA}^2] \\
&\quad + (i/\sigma^2)[\sigma_A^2 + (b^2)\tfrac{1}{2}\sigma_{AA}^2] + \ldots + (i/\sigma^2)[\sigma_A^2 + (b^{n-1})\sigma_{AA}^2].
\end{aligned}$$

This is a more general expression for μ_n than that previously obtained (Griffing 1960), when it was assumed that $y_f = y_m = y$. In this case, $b = (1-y)$.

(c) *Consequences of Relaxation after* n *Generations of Continuous Selection*

The objective in this section is to develop the prediction equation for the mean of a population which has had a history of n consecutive cycles of selection followed by t generations of random mating without selection. The procedure will be to start with Π_n and consider the consequences of one generation of relaxation, then to briefly outline the consequences of an arbitrary number t of generations of random mating without selection.

The notation is necessarily more complicated; thus $({}_f f_{ik}^{n,m})$ represents the frequency of the gamete $(A_i^1 A_k^2)$ produced by the females selected from the population $\Pi_{n,m-1}$ which has been subjected to n generations of continuous selection followed by $(m-1)$ generations of random mating without selection.

To obtain the mean of $\Pi_{n,1}$ it is necessary to start with

$$\Pi_{n,0} = \sum_{ijkl} ({}_f f_{ik}^{n,0})({}_m f_{jl}^{n,0})(A_i^1 A_k^2)(A_j^1 A_l^2).$$

The female genotype $(A_i^1 A_k^2)(A_j^1 A_l^2)$ produces the following gametic array

$$\{[(1-y_f)/2](A_i^1 A_k^2 + A_j^1 A_l^2) + (y_f/2)(A_i^1 A_l^2 + A_j^1 A_k^2)\}.$$

The total gametic array for the female population in which there is no selection is

$$\sum_{ijkl} ({}_f f_{ik}^{n,0})({}_m f_{jl}^{n,0})\{[(1-y_f)/2](A_i^1 A_k^2 + A_j^1 A_l^2) + (y_f/2)(A_i^1 A_l^2 + A_j^1 A_k^2)\} = \sum_{ik} ({}_f f_{ik}^{n,1})(A_i^1 A_k^2),$$

where it can be shown that

$${}_f f_{ik}^{n,1} = p_i^1 p_k^2 + n(i/\sigma^2) p_i^1 p_k^2 (\alpha_i^1 + \alpha_k^2) + (i/\sigma^2)(1-y_f)[(1-b^n)/(1-b)] p_i^1 p_k^2 (\alpha\alpha)_{ik}.$$

Likewise, the male frequency for the gamete $(A_j^1 A_l^2)$ is

$${}_m f_{jl}^{n,1} = p_j^1 p_l^2 + n(i/\sigma^2) p_j^1 p_l^2 (\alpha_j^1 + \alpha_l^2) + (i/\sigma^2)(1-y_m)[(1-b^n)/(1-b)] p_j^1 p_l^2 (\alpha\alpha)_{jl}.$$

Therefore, the mean of $\Pi_{n,1}$ is

$$\begin{aligned}\mu_{n,1} &\cong \sum_{ijkl} ({}_f f_{ik}^{n,1})({}_m f_{jl}^{n,1}) d_{ik.jl}^0 \\ &= (i/\sigma^2)[\sigma_A^2 + (b)\tfrac{1}{2}\sigma_{AA}^2] + (i/\sigma^2)[\sigma_A^2 + (b)(b)\tfrac{1}{2}\sigma_{AA}^2] \\ &\quad + (i/\sigma^2)[\sigma_A^2 + (b)(b^2)\tfrac{1}{2}\sigma_{AA}^2] + \ldots + (i/\sigma^2)[\sigma_A^2 + (b)(b^{n-1})\tfrac{1}{2}\sigma_{AA}^2].\end{aligned}$$

By working through successive cycles, it is clear that after t generations of random mating without selection, the female and male gametic frequencies are:

$${}_f f_{ik}^{n,t} = p_i^1 p_k^2 + n(i/\sigma^2) p_i^1 p_k^2 (\alpha_i^1 + \alpha_k^2) + (i/\sigma^2)(1-y_f)\{(b^{t-1})[(1-b^n)/(1-b)]\} p_i^1 p_k^2 (\alpha\alpha)_{ik},$$

and

$${}_m f_{jl}^{n,t} = p_j^1 p_l^2 + n(i/\sigma^2) p_j^1 p_l^2 (\alpha_j^1 + \alpha_l^2) + (i/\sigma^2)(1-y_m)\{(b^{t-1})[(1-b^n)/(1-b)]\} p_j^1 p_l^2 (\alpha\alpha)_{jl}.$$

Hence, the mean of the population which has been subjected to n cycles of continuous selection followed by t generations of random mating without selection is

$$\begin{aligned}\mu_{n,t} &\cong \sum_{ijkl}(_{f}f_{ik}^{n,t})(_{m}f_{jl}^{n,t})d_{ik.jl}^{0}\\ &= (i/\sigma^2)n\sigma_A^2+(i/\sigma^2)\{(b^t)[(1-b^n)/(1-b)]\}\tfrac{1}{2}\sigma_{AA}^2\\ &= (i/\sigma^2)n\sigma_A^2+(i/\sigma^2)[(b^t)(\sum_{r=1}^{n}b^{r-1})]\tfrac{1}{2}\sigma_{AA}^2\\ &= (i/\sigma^2)[\sigma_A^2+(b^t)\tfrac{1}{2}\sigma_{AA}^2]+(i/\sigma^2)[\sigma_A^2+(b^t)(b)\tfrac{1}{2}\sigma_{AA}^2]\\ &\quad + \ldots + (i/\sigma^2)[\sigma_A^2+(b^t)(b^{n-1})\tfrac{1}{2}\sigma_{AA}^2],\end{aligned}$$

where, as before,

$$b = \{[(1-y_f)/2]+[(1-y_m)/2]\} = [1-(y_f+y_m)/2].$$

Again, this is a generalization of the previous result when it was assumed that $y_f = y_m = y$.

If the increment change in means for the $(t-1)$th and tth populations is defined as

$${}_{n}\Delta\mu_{(t-1),t} = \mu_{n,(t-1)}-\mu_{n,t},$$

then

$$\begin{aligned}{}_{n}\Delta\mu_{(t-1),t} &= (i/\sigma^2)\{b^{t-1}[(1-b^n)/(1-b)]-b^t[(1-b^n)/(1-b)]\}\tfrac{1}{2}\sigma_{AA}^2\\ &= (i/\sigma^2)[b^{t-1}(1-b^n)]\tfrac{1}{2}\sigma_{AA}^2.\end{aligned}$$

For any value of n

$$\lim_{t\to\infty}[{}_{n}\Delta\mu_{(t-1),t}] \to 0,$$

hence the mean of the population $\Pi_{n,0}$ decays to

$$\lim_{t\to\infty}(\mu_{n,t}) \to (i/\sigma^2)n\sigma_A^2.$$

Finally, it must be noted that in all of the analyses of this and the previous section, it is assumed that natural selection is not operating in any way to modify the pressure applied by artificial selection.

III. Adapting the Two-locus Model to Genetically more Complex Situations

In the previous section the consequences of selection and relaxation from selection have been treated in detail for a very general genetic situation involving two loci. Clearly, however, if this form of analysis is to be of interest in selection theory, a method must be devised to adapt the two-locus model to accommodate, approximately, the variability generated by a much more complex genetic situation. This can only be done with certain simplifying assumptions. Thus, for the total genotypic variance, it is assumed that all epistatic interactions involving three

or more loci are negligible. For covariances which are disturbed by linkage, and for equations derived in the selection theory, it is assumed that: (1) all epistatic interactions involving three or more loci are negligible, and (2) the value for the covariance can be approximated by replacing the recombination value for the specific two-locus model by the recombination value averaged over all pairs of loci. The critical argument in this procedure is the argument on which the estimation of the "average recombination value" is based.

(a) *Average Recombination Value*

The problem in this section is to obtain an estimate of the recombination value averaged over all possible pairs of those loci whose genes cause variability in the characteristic under consideration.

The problem is particularly simple if all active loci are independent of each other in the segregational sense. In this case the average recombination value is, obviously, $\frac{1}{2}$. However, this situation implies that the individual loci are on different chromosomes, or, if two or more loci are on the same chromosome, they are spaced sufficiently far apart so that they segregate independently. Such restrictions make this simplified model unrealistic for most polygenic systems.

It is assumed in the following argument that for a given complexly inherited phenomenon, the number of active loci greatly exceeds the number of chromosome pairs, and that the genes are scattered over the chromosome complement. Hence, an entire range of linkage values among different pairs of loci is expected to occur. In this case it is obvious that the average recombination value over all pairs of loci lies between some lower limit and $\frac{1}{2}$.

In the first part of the following discussion, it will be shown that a lower limit can be determined quite simply using the chromosome as a unit of segregation. This argument will be followed by an attempt to bracket the average recombination value more exactly by taking crossing over into consideration. Finally it will be argued that a simple formula based only on the recombination index, although slightly biased, yields a sufficiently accurate estimate of the average recombination value for all practical purposes.

(b) *The Chromosome Argument*

For simplicity, consider the argument, first, for the situation of only two pairs of chromosomes. Assume that there are n_1 loci on one pair and n_2 loci on the other pair, where $n_1+n_2 = N$. The total number of different pairwise combinations of loci is

$$\binom{N}{2} = \frac{N(N-1)}{2}.$$

Each of these combinations falls into one of three classes. These classes are:

(i) Combinations in which both loci are located on the first chromosome. The number of such pairs is

$$\binom{n_1}{2} = \frac{n_1(n_1-1)}{2}.$$

(ii) Combinations in which both loci are located on the second chromosome. The number of such pairs is

$$\binom{n_2}{2} = \frac{n_2(n_2-1)}{2}.$$

(iii) Combinations of loci, one of which is located on the first chromosome and the other on the second chromosome. There are $n_1 n_2$ such combinations.

Each pair of loci in classes (i) and (ii) may exhibit a recombination value between 0 and $\frac{1}{2}$. However, it is assumed that each and every pair of loci in class (iii) exhibits a recombination value of exactly $\frac{1}{2}$. Hence, the lowest possible limit for the recombination value averaged over all possible pairs of loci is obtained by setting all recombination values in classes (i) and (ii) equal to zero. This lowest value is then

$$\tfrac{1}{2}[(n_1 n_2)/\tfrac{1}{2}N(N-1)] = n_1 n_2/N(N-1).$$

If $n_1 \cong N/2$, then the lowest average value becomes

$$N^2/4(N^2-N) = 1/4[1-(1/N)],$$

and the limit of this value as N becomes large is $\frac{1}{4}$.

This argument can be generalized easily to any number, m, of non-homologous chromosome sets. Suppose that there are n_i loci on the ith chromosome set ($i = 1, \ldots, m$) such that $N = \Sigma n_i$. The total number of different pairs of loci is

$$\binom{N}{2} = \frac{N(N-1)}{2}.$$

The number of pairs of loci, one on each of two non-homologous chromosomes is

$$\sum_{i<j} n_i n_j, \qquad (i,j = 1, \ldots, m).$$

Hence, the lowest possible average recombination value is

$$\sum_{i<j} n_i n_j/N(N-1). \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

If $n_i \cong (1/m)N$ (for all i), then (1) becomes

$$[(m-1)/2m][N^2/(N^2-N)] = [(m-1)/2m]\{1/[1-(1/N)]\}.$$

and the limit of this value as N becomes large is $(m-1)/2m$.

The above argument has been given in terms of chromosomes, and the lower limit of the average recombination value has been derived on the basis of no crossing over. In this case meiosis results in the independent segregation of entire chromosomes whose loci are completely linked. If crossing over occurs, it is possible to state that the recombination value averaged over all possible pairs of active loci lies in the interval

$$(m-1)/2m < \bar{y} < \tfrac{1}{2}.$$

It is instructive to tabulate $(m-1)/2m$ for varying values of m as follows (m = number of chromosome pairs):

m	1	2	3	4	5	6	10	20
$(m-1)/2m$	0	1/4	1/3	3/8	2/5	5/12	9/20	19/40

An interesting fact immediately becomes clear from this table, viz. the recombination value averaged over all possible pairs of loci is close to $\frac{1}{2}$ if the haploid chromosome number is five or more. This is due to the fact that as the number of chromosomes increases, the relative proportion of linked pairs of loci rapidly decreases and the average recombination value asymptotes steeply towards $\frac{1}{2}$.

However, there are undoubtedly many instances when attention is focused on sets of chromosomes whose numbers are small (i.e. less than five). This is particularly true for *Drosophila melanogaster*, since it is extensively used as an experimental organism in testing quantitative inheritance and selection theories. Therefore, it is of interest to extend the above argument in some detail.

(c) *The Crossing Over Argument*

The estimation of the average recombination value can be made more exact by considering the consequences of crossing over. This phenomenon increases the number of segregating units, and therefore, it is natural to suppose that an estimate of the average recombination value can be obtained by replacing m in the "chromosome" formula by Darlington's (1958) recombination index. This index gives the average number of pieces into which the chromosome complement is divided by chiasma formation.

However, this procedure, which may be termed the index method of estimation, is biased. This is so because the actual pattern of chiasma formation yields an extensive array of chromosome segments having different numbers of active loci, rather than a constant pattern of segments all having approximately the same number of loci. It will now be shown that, because of this fact, the index method yields an upper limit to the average recombination value. The argument will be illustrated first, for an obligatory chiasma forming on a single chromosome pair.

Suppose that a single chiasma forms at exactly the same position on the given chromosome pair in every mother cell. There will be n_1 active loci to the left and n_2 active loci to the right of the exchange. Clearly, pairs of loci to the left exhibit a recombination value of zero; as do pairs of loci to the right of the chiasma position. However, pairs of loci, which involve one locus on each side of the chiasma, exhibit a recombination value of $\frac{1}{2}$. Hence, the recombination value averaged over all loci is

$$\bar{y} = \tfrac{1}{2}n_1n_2 \Big/ \binom{N}{2}$$

$$= n_1n_2/N(N-1).$$

It can be shown that the average recombination value, $\bar{y}$, varies depending on the position of the chiasma, which in turn alters the relative magnitudes of n_1 and n_2. Furthermore, it can be shown that the maximum value of $\bar{y}$ is obtained for the situation in which $n_1 = n_2$. The argument may be sketched as follows: Let

$$N = n_1+n_2,$$

$$n_1 = n_2(1+\delta),$$

where $\delta > -1$. Then

$$n_1 = N[(1+\delta)/(2+\delta)],$$

and

$$n_2 = N/(2+\delta).$$

Hence

$$\begin{aligned}\bar{y} &= [(1+\delta)/(2+\delta)^2]\{1/[1-(1/N)]\} \\ &\cong (1+\delta)/(2+\delta)^2, \quad \text{for large } N, \\ &= 1/\{4+[\delta^2/(1+\delta)]\}.\end{aligned}$$

Since $\delta > -1$, $\delta^2/(1+\delta) \geqslant 0$. Therefore, the maximum value of $\bar{y}$ is $\frac{1}{4}$, which occurs when $\delta = 0$. This is the value obtained by the index estimation method.

This proof can be extended to the situation of more than one obligatory chiasma. The following argument for two chiasmata illustrates how this may be done.

Consider a hypothetical situation in which two chiasmata invariably form a given pattern in every mother cell. Let there be n_1 active loci to the left of the first chiasma, n_2 loci between the two chiasmata, and n_3 loci to the right of the second chiasma. Assuming no chromatid interference, the average recombination value over all pairs of loci is then,

$$\begin{aligned}\bar{y} &= \tfrac{1}{2}(n_1n_2+n_1n_3+n_2n_3)\Big/\binom{N}{2} \\ &= (n_1n_2+n_1n_3+n_2n_3)/N(N-1).\end{aligned}$$

The maximum value for $\bar{y}$ can be determined as follows:

Let

$$N = n_1+n_2+n_3,$$

$$n_1 = n_2(1+\alpha),$$

where $\alpha > -1$, and

$$n_3 = n_2(1+\beta),$$

where $\beta > -1$. Then

$$n_1 = [(1+\alpha)/(3+\alpha+\beta)]N, \quad n_2 = N/(3+\alpha+\beta), \quad \text{and} \quad n_3 = [(1+\beta)/(3+\alpha+\beta)]N.$$

Hence

$$\bar{y} = \frac{3+2\alpha+2\beta+\alpha\beta}{(3+\alpha+\beta)^2}\left[\frac{1}{1-(1/N)}\right],$$

which, for large N, is approximately

$$1\Big/\left[3+\frac{\alpha^2+\beta^2-\alpha\beta}{3+2\alpha+2\beta+\alpha\beta}\right].$$

Since $\alpha, \beta > -1$,

$$(\alpha^2+\beta^2-\alpha\beta)/(3+2\alpha+2\beta+\alpha\beta) \geqslant 0.$$

Thus, $\bar{y}$ is a maximum when

$$(\alpha^2+\beta^2-\alpha\beta)/(3+2\alpha+2\beta+\alpha\beta) = 0,$$

and this occurs only when both α and β equal zero. Therefore, the maximum value occurs when, invariably, $n_1 = n_2 = n_3$: the situation required for the argument involving the recombination index.

It is clear, then, that any agency which causes the positions of the chiasmata to be varied so that the chromosome pieces do not have equal contents of active loci lowers the average recombination value.

In reality, of course, the pattern of chiasmata is not invariable. Generally, chiasmata may form along the entire length of the chromosome, and the number of chiasmata for any given chromosome pair may vary in different mother cells. Observational data on the distribution of chiasmata in *individual* chromosomes are few, perhaps the most extensive are those reported by White and Morley (1955). However, it is apparently agreed that the following two conditions hold for most species of plants and animals:

(1) At least one chiasma per bivalent is obligatory for the survival of the bivalent; and

(2) A strong chiasma interference exists, at least within each arm of every chromosome.

Since there is no chiasma distributional theory which completely satisfies the above conditions, the procedure which will be followed is to continue the approach of using simplified cross-over models to bracket the true recombination value in as small an interval as possible. Finally, it will be shown that the index method, although biased, yields a sufficiently accurate estimate for most practical situations.

In view of the fact that the pattern of chiasmata is not invariable, it is clear that the index method yields an upper limit to the average recombination value. Therefore the true average recombination value must lie in the interval

$$(m-1)/2m < \bar{y} < (r-1)/2r,$$

where m = haploid chromosome number, and r = recombination index.

In this interval, the upper limit is set by a cross-over model in which the chiasma configuration invariably yields chromosomal segments containing equal numbers of active loci. It is "sensitive" in that the limit changes with different numbers of chiasmata for a given number of chromosomes. The lower limit, however, is based on chromosomal segregation and is "insensitive" to chiasma distribution. Hence the next step is to devise a cross-over model which yields a sensitive lower limit to the true average recombination value.

As mentioned earlier, one of the accepted facts with regard to actual chiasma distribution is the strong chiasma interference within a chromosome arm. Such interference tends to space the chiasmata located in the same arm. This results in a restricted array of chromosomal segments tending to have the same number

of active loci: the condition required by the index method. Hence, a cross-over model in which there is no chiasma interference will produce an unrestricted array of chromosome segments, with the net result that the average recombination value based on this model will generally be lower than the true value. Thus, a non-interference model, incorporating the following conditions, will be considered as providing a sensitive lower limit for the true average recombination value:

(1) At least one obligatory chiasma per bivalent;

(2) Non-interference of chiasma formation, i.e. all chiasma form independently of each other; and

(3) No chromatid interference.

In determining the consequences of this non-interference model, it is convenient to break down the approach into two stages, both of which are concerned with chiasma formation on a *single chromosome pair*. In stage one, the average recombination value is determined for a given number, k, of independent chiasmata formed on the single chromosome pair in every mother cell. In stage two, the numbers of chiasmata are allowed to vary according to the Poisson distribution.

With regard to stage one, first consider a specific example in which there are three independent chiasmata formed on the given pair of chromosomes in every mother cell. Let the length of the chromosome be divided by n loci into $(n-1)$ regions, in each of which a chiasma is equally likely to occur. (It is assumed that the regions are sufficiently small so that the probability of two chiasmata forming in the same region is negligible).

There are $\binom{n}{2}$ different pairwise combinations of loci, and $\binom{n-1}{3}$ different chiasma configurations. Hence, there is a total of $\binom{n}{2}\binom{n-1}{3}$ events, which, for a given pair of loci, may be defined in terms of the recombinational consequences of the imposition of a certain chiasma configuration. Thus a recombinant event is one which, when all possible meiotic configurations are considered, results in 50 per cent. recombinant chromosomes. Such an event occurs when at least one chiasma forms between the two loci, and a non-recombinant event occurs when chiasmata do not form between the loci.

The recombination value averaged over all possible pairs of loci may then be defined as

$$\bar{y} = \left(\frac{\text{Number of recombinant events}}{\text{Total number of events}}\right) \times \tfrac{1}{2}.$$

This is most easily obtained as

$$\bar{y} = \left(\frac{(\text{Total number of events}) - (\text{Number of non-recombinant events})}{\text{Total number of events}}\right) \times \tfrac{1}{2}.$$

The basic problem, then, is the enumeration of the non-recombinant events.

For a given chromosome configuration, a non-recombinant event exists when the region between the two loci is not interrupted by one or more chiasmata. Hence, for the three-chiasmata example, the enumeration problem is simplified if, for each chiasma configuration, the number of combinations of loci which occur, (1) to the left of the first chiasma, (2) between the first and second chiasmata, (3) between the second and third chiasmata, and (4) to the right of the third chiasma, are enumerated. The total number of such non-recombinants is found to be

$$4\sum_{r=2}^{n-3}\binom{n-r-1}{2}\binom{r}{2}.$$

Therefore, the average recombination value, as n increases indefinitely, for a single chromosome pair invariably having three chiasmata is

$$\bar{y} = \lim_{n\to\infty}\left[\frac{\binom{n}{2}\binom{n-1}{3}-4\sum_{r=2}^{n-3}\binom{n-r-1}{2}\binom{r}{2}}{\binom{n}{2}\binom{n-1}{3}}\right]\times\tfrac{1}{2}$$

$$= \lim_{n\to\infty}\left[\frac{(1/12)n(n-1)^2(n-2)(n-3)-(1/30)n(n-1)(n-2)(n-3)(n-4)}{(1/12)n(n-1)^2(n-2)(n-3)}\right]\times\tfrac{1}{2}$$

$$= \lim_{n\to\infty}\left\{3/10\left[\frac{1+(1/n)}{1-(1/n)}\right]\right\}$$

$$= 3/10.$$

This argument can be readily generalized to any number, k, of independent chiasmata formed on the given pair of homologous chromosomes in each mother cell. This generalization results in the following expressions for the average recombination value:

$$\bar{y} = \lim_{n\to\infty}\left\{\frac{\binom{n}{2}\binom{n-1}{k}-(k+1)\sum_{r=2}^{n-k}\binom{n-r-1}{k-1}\binom{r}{2}}{\binom{n}{2}\binom{n-1}{k}}\right\}\times\tfrac{1}{2}$$

which is, apparently,

$$= \lim_{n\to\infty}\left\{\frac{\frac{1}{2k!}\left[\frac{n(n-1)(n-1)!}{(n-k-1)!}\right]-\left[\frac{n!}{k!(k+2)(n-k-2)!}\right]}{\frac{1}{2k!}\left[\frac{n(n-1)(n-1)!}{(n-k-1)!}\right]}\right\}\times\tfrac{1}{2}$$

$$= \lim_{n\to\infty}\left\{\left(\frac{k}{k+2}\right)\left(\frac{1+(1/n)}{1-(1/n)}\right)\right\}\times\tfrac{1}{2}$$

$$= k/2(k+2).$$

The above argument is developed for the genetic situation in which exactly k independent chiasmata occur on the given bivalent in every mother cell. It is now necessary to extend this argument to permit varying numbers of chiasmata to form on the chromosome pair in different mother cells. The extension must satisfy the two conditions mentioned previously, i.e. (1) at least one chiasma is obligatory for each bivalent, and (2) all chiasmata form independently of each other (no chiasma interference). It follows from the last condition that, apart from the initial obligatory chiasma, the numbers of additional chiasmata are distributed according to the Poisson distribution. Thus, the probability of exactly t additional chiasmata is

$$P(t;\lambda) = \frac{e^{-\lambda}\lambda^t}{t!},$$

where λ = (average number of chiasmata)-1.

The frequencies for varying t values are as follows:

t	0	1	2	3	...
Obligatory chiasma	1	1	1	1	...
Total chiasmata ($k = t+1$)	1	2	3	4	...
Frequency	$e^{-\lambda}$	$e^{-\lambda}\lambda$	$(e^{-\lambda}\lambda^2)/2!$	$(e^{-\lambda}\lambda^3)/3!$	...

The average recombination value for *exactly k chiasmata in each and every cell* may be recast in terms of the variable t as follows:

$$\bar{y} = \frac{k}{2(k+2)}$$

$$= \frac{(t+1)}{2(t+3)}.$$

The average recombination value over all possible pairs of loci may now be obtained for the situation in which the frequencies for varying values of t are taken into consideration as follows:

$$\bar{y} = \sum_{t=0}^{\infty}[(e^{-\lambda}\lambda^t)/t!][(t+1)/2(t+3)]$$

$$= \sum_{t=0}^{\infty}[(e^{-\lambda}\lambda^t)/t!][\tfrac{1}{2}-1/(t+3)]$$

$$= \tfrac{1}{2}\sum_{t=0}^{\infty}[(e^{-\lambda}\lambda^t)/t!]-(1/e^{\lambda})\{\sum_{t=0}^{\infty}[\lambda^t/t!(t+3)]\}$$

$$= \tfrac{1}{2}-(1/e^{\lambda})\{(2/\lambda^3)[(\lambda^2 e^{\lambda}/2)-(\lambda-1)e^{-\lambda}-1]\}$$

$$= \tfrac{1}{2}-(1/\lambda)+[2(\lambda-1)/\lambda^3]+(2/\lambda^3 e^{\lambda}).$$

The value $\bar{y}$ then is the recombination value averaged over all possible pairs of loci for a single chromosome pair when it is assumed that:

(1) a large number of active loci occur at random in the chromosome pair;

(2) at least one chiasma is obligatory for the survival of the bivalent; and

(3) there is neither chiasma nor chromatid interference.

The above expression holds for $\lambda \geqslant 1$; for $\lambda = 0$, $\bar{y} = \frac{1}{6}$.

To summarize the results when crossing over is taken into consideration, the true average recombination value for a single chromosome pair lies in the interval defined

TABLE 1

DIFFERENCE OF THE AVERAGE RECOMBINATION VALUE FOR THE METHOD IN WHICH THE RECOMBINATION INDEX IS USED AND THE METHOD IN WHICH THE CHIASMATA ARE ASSUMED TO FORM INDEPENDENTLY OF EACH OTHER WITH THE RESTRICTION THAT AT LEAST ONE CHIASMA IS OBLIGATORY

Average No. of Chiasmata per Chromosome (s)	No. of Sets of Homologous Chromosomes (m)				
	1	2	3	4	5
1	0·083	0·042	0·028	0·021	0·017
2	0·098	0·049	0·033	0·024	0·020
3	0·091	0·046	0·030	0·023	0·018
4	0·081	0·041	0·027	0·020	0·016
5	0·072	0·036	0·024	0·018	0·014

by the non-interference model at the lower limit and the index model at the upper limit, i.e.

(1) Single obligatory chiasma:

$$\tfrac{1}{6} < \bar{y} < \tfrac{1}{4}.$$

(2) Single obligatory chiasma plus λ additional chiasmata:

$$\{\tfrac{1}{2} - (1/\lambda) + [2(\lambda-1)/\lambda^3] + (2/\lambda^3 e^{\lambda})\} < \bar{y} < [(\lambda+1)/2(\lambda+2)].$$

The magnitudes of the intervals for different average numbers of chiasmata are given in the first column of Table 1. These values vary from 7 to 10 per cent. and tend to diminish as the average number of chiasmata increases. Because (i) it is likely that the probability distribution of chiasma position is not uniform throughout a chromosome arm, and (ii) a powerful chiasma interference occurs which tends to disperse the points of exchange, the true average recombination value, in most instances, will probably lie toward the upper limit of the interval, not far from the value given by the index method. For a single chromosome, then, the index method yields an estimate with a slight positive bias.

In selection theory, however, one is seldom concerned with a single chromosome; therefore, it is of interest to determine the magnitude of the interval generated by the non-interference (independent chiasmata) and the index methods, as the chromosome number increases. The following gives the details for (i) m chromosome pairs each having only one chiasma, and (ii) m chromosome pairs each having an average of $s = (\lambda+1)$ chiasmata.

For the situation in which each of m chromosome pairs has only one chiasma, the two models yield the following formulae:

$$\bar{y}_{\text{index}} = (2m-1)/4m,$$

and

$$\bar{y}_{\text{indep.}} = (3m-2)/6m.$$

Hence, the difference between the two estimators is

$$d = \bar{y}_{\text{index}} - \bar{y}_{\text{indep.}} = 1/12m.$$

For the situation in which each of m chromosome pairs has an average of $s = (\lambda+1)$ chiasmata, the corresponding formulae are

$$\bar{y}_{\text{index}} = [m(\lambda+2)-1]/[2m(\lambda+2)],$$

$$\bar{y}_{\text{indep.}} = \tfrac{1}{2} - (1/m\lambda) + [2(\lambda-1)/m\lambda^3] + 2e^{-\lambda}/m\lambda^3,$$

and

$$d = \bar{y}_{\text{index}} - \bar{y}_{\text{indep.}} = \{(\lambda^3-4\lambda+8)/[2m(\lambda+2)\lambda^3]\} - (2e^{-\lambda}/m\lambda^3).$$

The magnitudes of the differences for varying values of m and s are given in Table 1. It is clear that as the number of chromosomes increases, the differences generated by the two methods of estimation diminish, and both methods converge on the true value.

It may be concluded that for all practical purposes the index method, which is simple and convenient to use, yields a satisfactory estimate of the average recombination value. For example, when the chromosome number is increased to only two, the true recombination value may be expected to lie within 2 per cent. of the value given by the index method. If greater precision is required, estimates could be made using both methods and the arithmetic mean obtained from them.

The methods outlined in this section give at least a first approximation to the recombination value averaged over all pairs of loci. The approximation becomes better as more information on chiasma frequency becomes available. If only the chromosome number is known, one can certainly say that the average recombination value lies in the range

$$(m-1)/2m < \bar{y} < \tfrac{1}{2},$$

where m is the haploid chromosome number. If fertility is high, it can be assumed that at least one obligatory chiasma occurs in each and every chromosome pair. Hence, without actually making a chiasma count, one can obtain an estimate of the

average recombination value, which should be fairly accurate in most cases, by the following formula

$$\bar{y} = (m'-1)/2m',$$

where m' is twice the haploid chromosome number.

Finally, if the average number of chiasmata per cell is known, it is possible to bracket the average recombination value in a smaller range, i.e.

$$\left\{\tfrac{1}{2} - \frac{1}{p-m} + \frac{2m^2}{(p-m)^3}\left[\exp\left(-\frac{p-m}{m}\right) + \frac{p-2m}{m}\right]\right\} < \bar{y} < \frac{p+m-1}{2(p+m)},$$

where the expression to the left is the average recombination value for the non-interference model when there is an average of $(\lambda+1) = p/m$ chiasmata on each of m chromosome pairs, and p is the average number of chiasmata per nucleus. When the number of chiasmata vary with different chromosomes, this expression is not exact. The expression on the right is the average recombination value given by the index method.

For practical purposes a direct estimate of the average recombination value may be made, using the convenient and simple index formula, i.e.

$$\bar{y} = (p+m-1)/2(p+m).$$

The above argument supposes that (i) there are a very large number of active loci scattered at random over the chromosome set, (ii) the chromosomes are not drastically different in size (as, for example, with the two major and the fourth autosomes in *Drosophila*), and (iii) a chiasma is invariably associated with a genetic crossing over.

(*d*) *A* Drosophila *Example*

Clayton, Morris, and Robertson (1957) and Clayton and Robertson (1957) demonstrated that with the population of *Drosophila* which they were using, almost all of the genetic variability for abdominal bristle number was generated by genes located in the two major autosomes. The question may now be asked as to what is the average recombination value for each of the two sexes for all possible pairs of genes causing variability in bristle number.

Since crossing over does not occur in the male, the recombination index for these autosomes is merely 2, and hence the average recombination value for males is equal to $\frac{1}{4}$.

Crossing over occurs in the female, enabling linkage maps to be synthesized. Since the map for each of the two major autosomes is slightly greater than 100 cross-over units, the recombination index for the two major autosomes is approximately 6. Therefore, the average recombination value for females is approximately 5/12.

IV. Estimates of Variance Components from Covariances when the Recombination Values are Different in the Two Sexes

In a previous section formulae were given for the responses to selection and relaxation from selection in terms of genotypic variance components. In comparing

the theoretical consequences of selection with the experimental results, it is necessary to estimate these variance components from the original random-mating population. This is done with covariances.

The expectations of some of these covariances are affected by linkage, and Cockerham (1956) has given the expectations of the necessary covariances for the situation in which the recombination value is the same for both sexes. These expressions must now be generalized to accommodate different recombination values for the two sexes.

There are three covariances which are of interest. These are: parent–offspring covariance, designated as Cov(PO); half-sib covariance, designated as Cov(HS); and full-sib covariance, designated as Cov(FS). Of these, the Cov(PO) is not affected by linkage, but linkage parameters do enter into the expectations for the other two covariances.

The definition of the half-sib covariance may be reduced to the expectation of the squares of the half-sib family means. A "sire" half-sib family is generated by the union of the gametic array of an arbitrary sire with the total gametic array from the dams, and, similarly, a "dam" half-sib family results from the union of the gametic array of an arbitrary dam with the total gametic array from the sires. If the recombination value is different for the two sexes, it is obvious that the sire and dam half-sib families differ, even if the sire and dam are of the same genotype. Thus, two different covariances are possible:

$\mathrm{Cov}_{(m)}$ (HS) = covariance generated by sire half-sib families, and

$\mathrm{Cov}_{(f)}$ (HS) = covariance generated by dam half-sib families.

Consider first, the derivation of $\mathrm{Cov}_{(m)}$ (HS). An arbitrary sire $(A_i^1A_k^2)(A_j^1A_l^2)$, produces the following gametic array

$$\{[(1-y_m)/2](A_i^1A_k^2+A_j^1A_l^2)+(y_m/2)(A_i^1A_l^2+A_j^1A_k^2)\}.$$

The total female gametic array in the random-mating population is

$$\sum_{rt} p_r^1 p_t^2 (A_r^1 A_t^2).$$

Hence, the sire half-sib family mean is

$$h_{(ik.jl)(....)} = [(1-y_m)/2][\sum_{rt} p_r^1 p_t^2 d_{ik.rt}+\sum_{rt} p_r^1 p_t^2 d_{jl.rt}]$$
$$+(y_m/2)[\sum_{rt} p_r^1 p_t^2 d_{il.rt}+\sum_{rt} p_r^1 p_t^2 d_{jk.rt}].$$

The sire half-sib covariance may then be evaluated as

$$\mathrm{Cov}_{(m)}\ (\mathrm{HS}) = \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 [h_{(ik.jl)(....)}]^2$$
$$= \tfrac{1}{4}\sigma_A^2+[(1/16)+(\delta_m/16)]\sigma_{AA}^2,$$

where

$$\delta_m = (1-2y_m)^2.$$

Likewise, the dam half-sib covariance is

$$\mathrm{Cov}_{(f)}\ (\mathrm{HS}) = \tfrac{1}{4}\sigma_A^2+[(1/16)+(\delta_f/16)]\sigma_{AA}^2,$$

where

$$\delta_f = (1-2y_f)^2.$$

The defintion of the full-sib covariance may be reduced to the expected value of the squares of the full-sib means. Consider, now, the evaluation of this definition for the case of different recombination values for the two sexes.

The mean of the full-sib array which results from the cross between an arbitrary sire, $(A_i^1A_k^2)(A_j^1A_l^2)$, and an arbitrary dam, $(A_r^1A_t^2)(A_s^1A_u^2)$, is

$$\begin{aligned} h_{(ik.jl)(rt.su)} = \{&[(1-y_m)/2][(1-y_f)/2](d_{ik.rt}+d_{ik.su}+d_{jl.rt}+d_{jl.su}) \\ &+[(1-y_m)/2](y_f/2)(d_{ik.ru}+d_{ik.st}+d_{jl.ru}+d_{jl.st}) \\ &+(y_m/2)[(1-y_f)/2](d_{il.rt}+d_{il.su}+d_{jk.rt}+d_{jk.su}) \\ &+(y_m/2)(y_f/2)(d_{il.ru}+d_{il.st}+d_{jk.ru}+d_{jk.st})\}. \end{aligned}$$

By definition, the full-sib covariance may be obtained as follows:

$$\begin{aligned} \mathrm{Cov(FS)} &= \sum_{ijklrstu} p_i^1p_j^1p_r^1p_s^1p_k^2p_l^2p_t^2p_u^2[h_{(ik.jl)(rt.su)}]^2 \\ &= \tfrac{1}{2}\sigma_A^2+\tfrac{1}{4}\sigma_D^2+\{\tfrac{1}{4}+[(\delta_f+\delta_m)/16]\}\sigma_{AA}^2 \\ &\quad +\{\tfrac{1}{8}+[(\delta_f+\delta_m)/16]\}\sigma_{AD}^2+(1/16)(1+\delta_f)(1+\delta_m)\sigma_{DD}^2. \end{aligned}$$

Assuming that the epistatic interactions involving three or more loci are negligible, the simplest method of estimating σ_A^2 and σ_{AA}^2 is to use the parent–offspring and half-sib covariances. Thus, starting with the expectations

$$\mathrm{Cov\ (PO)} = \tfrac{1}{2}\sigma_A^2+\tfrac{1}{4}\sigma_{AA}^2,$$

and

$$\mathrm{Cov}_{(m)}\ (\mathrm{HS}) = \tfrac{1}{4}\sigma_A^2+1/16(1+\delta_m)\sigma_{AA}^2,$$

estimates of σ_A^2 and σ_{AA}^2 may be obtained as follows:

$$\hat{\sigma}_A^2 = \{\mathrm{Cov(PO)}[2y_m(1-y_m)-1]+2[\mathrm{Cov}_{(m)}(\mathrm{HS})]\}/y_m(1-y_m),$$

and

$$\hat{\sigma}_{AA}^2 = 2\{\mathrm{Cov(PO)}-2[\mathrm{Cov}_{(m)}(\mathrm{HS})]\}/y_m(1-y_m).$$

Normally, then, the experimental procedure would be to obtain accurate estimates of the parent–offspring and half-sib covariances from the original random-mating population, and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_{AA}^2$ using the above expressions in which the specific recombination values are replaced by the recombination value averaged over all possible pairs of active loci. The variance estimates may then be substituted into the theoretical selection formulae for comparison with the observed responses.

V. A Theoretical Method for the Detection of Disturbances due to Natural Selection

The selection theory has been developed on the basis that *natural* selection is not operating differentially on the various genotypes. Clearly, however, this need not be the case. It appears that in experiments where artificial selection has been applied unidirectionally, it has often been found that when the population is relaxed from selection, the mean regresses toward its original value. This phenomenon has been termed "genetic homeostasis" by Lerner (1954), and it is generally assumed that this regression is due to the effects of natural selection which oppose the effects of artificial selection.

However, in this study, and in a previous paper (Griffing 1960), it is shown that the contributions of certain epistatic effects mimic the antogonistic effects of natural selection. This mimicry occurs in both the response to selection and relaxation from selection. Thus, even when natural selection is ineffective, the increment changes due to artificial selection in successive generations become increasingly smaller: i.e. in the formula

$$\Delta\mu_{n,(n-1)} = (i/\sigma^2)[\sigma_A^2+(b^{n-1})\tfrac{1}{2}\sigma_{AA}^2],$$

the contribution of σ_{AA}^2 decreases as n increases. This diminishing of increments is similar to the effect that one would expect on the assumption that the intensity of natural selection increases as artificial selection causes the cumulative change in the population mean to increase.

Likewise, the decay of the mean on relaxation from the value

$$\mu_{n,0} = (i/\sigma^2)[n\sigma_A^2+\sum_{r=1}^{n}(b^{r-1})\tfrac{1}{2}\sigma_{AA}^2]$$

to

$$\lim_{t\to\infty}(\mu_{n,t}) \to (i/\sigma^2)n\sigma_A^2,$$

simulates the response which would occur if natural selection were operating in the absence of artificial selection to regress the mean toward its original unselected value.

Thus, if epistatic contributions are not taken into consideration, the disturbance they cause may be confounded with, or wrongly judged due to, the antagonistic effects of natural selection. Therefore, it is necessary to outline a possible method of detecting the influence of natural selection in the presence of disturbances caused by epistatic effects. Such a method will be given after the basic requirements for a selection programme aimed at detecting natural selection are given.

(a) *Requirements of a Selection Programme*

The following lists the basic requirement in a selection programme designed to detect the effects of natural selection in opposing artificial selection:

(i) Use of a truly random-mating population in equilibrium as the original population with which the selection programme starts;

(ii) Collection of sufficient data to allow for the accurate estimation of parent–offspring and half-sib covariances in the original population (this requires a very large sample of observations);

(iii) Collection of data from several cycles of truncation selection based on the individual phenotype (the selection must be conducted with known intensity and with large numbers of individuals); and

(iv) Collection of data from selected lines which have been allowed to mate at random for several generations, i.e. from relaxed lines.

There are other experimental ramifications which are useful. These include the following:

(v) Simultaneous selection experiments conducted with the same intensity in opposite directions;

(vi) Simultaneous selection experiments conducted with different intensities of selection;

(vii) Relaxation of selection at various stages in a continuous selection programme;

(viii) Back-selection of relaxed lines after they have become stabilized;

(ix) Mass reciprocal crossing of highly selected *relaxed* lines; and

(x) Sufficient fitness measurements to give a developmental picture of the change in fitness throughout the selection programme.

These subsidiary experimental procedures give additional information, allow a broader basis for the comparison of experimental results with selection theory, and provide some measure for the verification of the validity of the assumptions on which the theory is based. For example, simultaneous selection in opposite directions provides a useful check on the fulfilment of the assumptions, since if the assumptions hold, a symmetrical response should be obtained by the application of identical "up" and "down" selection pressures. If an asymmetrical pattern develops, at least one of the assumptions has been violated.

(*b*) *A Method for Detecting the Effects of Natural Selection*

When data from a selection programme, as outlined above, are available, the influence of natural selection may be detected, in the presence of epistatic disturbances, by comparing the observed responses due to artificial selection with those expected from the theory presented earlier which assumes no disturbance from natural selection. This requires the computation of accurate, unbiased estimates of σ^2_A and σ^2_{AA} from the original unselected population and replacing the variances in the selection formulae with these estimates. If the observed means do not differ significantly from the expected values, it may be inferred, generally, that natural selection does not have an appreciable effect. If, however, the observed means are significantly less than the expected values, it may generally be assumed that natural selection is opposing the effects of artificial selection. This argument holds only if the assumptions inherent in the theory have not been violated.

VI. Discussion

This paper deals primarily with the extension of mass selection theory to accommodate linkage. The various aspects of this problem are developed in three stages.

(i) Mass selection theory for two loci is extended to accommodate different recombination values for the two sexes. This theory is now completely general in that it permits: (1) any number of alleles at each locus; (2) arbitrary linkage, including the condition that the recombination value may be different for the two sexes; and (3) arbitrary dominance and epistatic effects.

This extension of the two-locus theory is necessary if one is to take into consideration the general phenomenon of linkage in organisms such as *Drosophila* which have different recombination values in the two sexes.

(ii) A method is developed for estimating the recombination value averaged over all possible pairs of loci scattered at random over all chromosome sets.

The solution to this problem is necessary in order to adapt the generalized two-locus theory, developed in the first stage, to cope with genetic situations which are considerably more complex. That is to say, the solution to stage (ii) allows the variability generated by a genetically complex system involving many loci to be approximately described by the relatively simple two-locus theory.

(iii) The expectations of the half-sib and full-sib covariances for a random-mating population are generalized to permit different recombination values for the two sexes.

Solution to stage (iii) is necessary for the estimation of genotypic variance components from covariances between relatives which are, themselves, subject to linkage disturbances. These variance components can then be used for the purpose of comparing observed with theoretical selection responses.

Finally, the above theory is used to outline a method which permits the detection of the influence of natural selection in modifying the effectiveness of artificial selection even when the mimicking effects of epistasis are present.

Clearly, the approach to the linkage problem adopted in this study results in only an approximate solution, since various simplifying assumptions are required. These include: (1) epistatic interactions involving three or more loci are negligible, and (2) the true value of $\mathrm{Cov}_{(m)}(\mathrm{HS})$ can be approximated by the $\mathrm{Cov}_{(m)}(\mathrm{HS})$ derived from the two-locus model in which the average recombination value is substituted for the specific value for the two loci.

The usefulness of the theory rests on its ability to describe observed selection results. At the present time, the only data available which are sufficiently comprehensive to make this comparison are those reported by Clayton, Morris, and Robertson (1957) and Clayton and Robertson (1957). Unfortunately, however, the basic covariances from the original population were not estimated with sufficient accuracy to provide meaningful estimates of the variance components. Therefore, comparison of observed responses with the theoretical results of the present study will have to be made in the future when sufficiently accurate data are available.

VII. Acknowledgments

I am grateful to my colleague, Dr. R. N. Oram, for his valuable suggestions regarding the crossing-over argument.

VIII. References

Clayton, G. A., Morris, J. A., and Robertson, A. (1957).—An experimental check on quantitative genetical theory. I. Short-term responses to selection. *J. Genet.* **55**: 131–51.

Clayton, G. A., and Robertson, A. (1957).—An experimental check on quantitative genetical theory. II. The long-term effects of selection. *J. Genet.* **55**: 152–70.

Cockerham, C. C. (1956).—Effects of linkage on the covariances between relatives. *Genetics* **41**: 138–41.

Darlington, C. D. (1958).—"Evolution of Genetic Systems." 2nd Ed. (Oliver and Boyd: London.)

Griffing, B. (1960).—Theoretical consequences of truncation selection based on the individual phenotype. *Aust. J. Biol. Sci.* **13**: 307–43.

Kempthorne, O. (1954).—The correlations between relatives in a random mating population. *Proc. Roy. Soc.* B **143**: 103–13.

Kempthorne, O. (1957).—"An Introduction to Genetic Statistics." (John Wiley and Sons, Inc.: New York.)

Kimura, M. (1958).—On the change of population fitness by natural selection. *Heredity* **12**:145–67.

Lerner, I. M. (1954).—"Genetic Homeostasis." (Oliver and Boyd: London.)

White, M. J. D., and Morley, F. H. W. (1955).—Effects of pericentric rearrangements on recombination in grasshopper chromosomes. *Genetics* **40**: 604–19.