SELECTION IN REFERENCE TO BIOLOGICAL GROUPS

II.* CONSEQUENCES OF SELECTION IN GROUPS OF ONE SIZE WHEN EVALUATED IN GROUPS OF A DIFFERENT SIZE†

By B. GRIFFING‡

[*Manuscript received June* 11, 1968]

Summary

Consequences of individual and group selection are examined for the case in which selection operates with regard to groups of one size and its effects are measured with regard to groups of a different size. In such a situation neither selection procedure can ensure that positive selection will result in a non-negative change in the population mean.

Selection among sets of randomly associated groups is less efficient than selection among individual groups. Likewise, it is shown that if individuals within groups are independent, group selection is less efficient than individual selection.

I. INTRODUCTION

In the first paper of this series (Griffing 1967)* the genetic model usually used in selection theory was extended to accommodate any form of interaction (cooperative or competitive) between genotypes within small groups. The consequences of individual and group selection procedures were derived with reference to this more complex, but yet more realistic, biological representation.

The results of this first study can be summarized as follows. The change in the population mean due to individual selection is a function of a sum of cross-products involving direct and associate additive effects of the genetic model. In certain instances, positive individual selection can result in a negative sum of cross-products which, in turn, can cause a negative change in the mean. Thus, the incongruous situation can arise in which continued positive individual selection results in a deterioration of the population genotypic structure ultimately leading to fixation of the least desirable allele.

The dilemma can be resolved by group selection, since the change in the population mean resulting from positive group selection is a function of a sum of squares rather than a sum of cross-products. Hence, the change in the population mean cannot be negative. Thus theoretically, transferring the basis of selection from that of the individual to that of the group ensures a desirable response in the population.

It is also pointed out in the discussion of the first paper that a study of the consequences of selection in the most general terms requires consideration of two

* Part I, *Aust. J. biol. Sci.*, 1967, **20**, 127–39.

† This work was supported by a grant from the United States Atomic Energy Commission [contract number AT(11–1)–1761].

‡ Faculty of Genetics, Ohio State University, Columbus, Ohio 43210.

populations of groups: (1) the population of groups in which selection operates, and (2) the population of groups in which the effects of selection are measured. In the analyses of the first paper, these two populations are assumed to be the same. However, there are situations in which this need not be the case. For example, in many plant-breeding applications, certain selection procedures are conducted among spaced plants under the assumption that the desired responses will be exhibited by progeny grown under sward conditions.

In the present study, results are obtained for individual and group selection operating with regard to groups of one size but evaluated with regard to groups of a different size. Furthermore, the consequences of selection are derived for a higher order of organization: that of sets (or constellations) of groups. In this case selection among individuals as well as selection among sets is considered. Thus the following question is examined: if it is possible to ensure a desirable response by shifting the basis of selection from that of the individual to that of the group, is it also possible to obtain greater response by extending the process so that selection operates on a "group" of groups rather than on a single group?

II. Consequences of Selection

(a) *Specification of the Model*

Specification of the total biological model which includes the definition of a group, the construction of a population of groups, and the identification of the associated gene model is given in detail in the first paper of this series. Therefore this aspect of the problem will be stated only briefly here.

Consider a base population of genotypes generated by an arbitrary number of alleles at a single locus. The genotypic array of this population, which is assumed to be in equilibrium under random mating, may be represented as follows:

$$\sum_{i,j} p_i p_j (A_i A_j).$$

The population of groups of size n is obtained from an n-way combinatorial product involving the base population, e.g.

$$[\sum p_i p_j (A_i A_j)] \times [\sum p_i p_j (A_i A_j)] \times \cdots \times [\sum p_i p_j (A_i A_j)]$$
$$= \sum p_{i_1} p_{j_1} p_{i_2} p_{j_2} \cdots p_{i_n} p_{j_n} (A_{i_1} A_{j_1}, A_{i_2} A_{j_2}, \ldots, A_{i_n} A_{j_n}).$$

In this representation each group of n individuals is characterized by an appropriate unordered n-tuple, i.e.

$$(A_{i_1} A_{j_1}, A_{i_2} A_{j_2}, \ldots, A_{i_n} A_{j_n}).$$

The genotypic value of $A_{i_1} A_{j_1}$ as expressed in the above n-tuple is denoted as

$$_{i_1 j_1} d_{i_2 j_2, \ldots, i_n j_n},$$

and coded so that

$$\sum p_{i_1} p_{j_1} \cdots p_{i_n} p_{j_n} ({}_{i_1 j_1} d_{i_2 j_2, \ldots, i_n j_n}) = 0.$$

The subscripts in front of the symbol d indicate the genetic constitution of the genotype under consideration, and the subscripts following d indicate the genetic constitution of the remaining $(n-1)$ associated genotypes in the group.

The gene model for any given genotype must not only represent the *direct* effects of its own genes but also the *associate* gene effects of other members in the group. Hence, although only one locus is involved, an n-locus model must be used to completely characterize the genotypic value of an individual in a group of size n. Thus the gene model for $A_{i_1}A_{j_1}$ as expressed in the n-tuple $(A_{i_1}A_{j_1}, \ldots, A_{i_n}A_{j_n})$ is

$$\begin{aligned} {}_{i_1j_1}d_{i_2j_2,\ \ldots,\ i_nj_n} = {}_d\alpha_{i_1} + {}_d\alpha_{j_1} + {}_d\delta_{i_1j_1} + {}_a\alpha_{i_2} + {}_a\alpha_{j_2} + {}_a\delta_{i_2j_2} \\ + \ldots + {}_a\alpha_{i_n} + {}_a\alpha_{j_n} + {}_a\delta_{i_nj_n} + \ldots + {}_{da}(\alpha\alpha)_{i_1i_2} + \ldots \end{aligned}$$

where

$$\begin{aligned} {}_d\alpha_{i_1} &= {}_{i_1}.d_{..,\_\_,..} = \sum p_{j_1}p_{i_2}p_{j_2} \ldots p_{i_n}p_{j_n}({}_{i_1j_1}d_{i_2j_2,\ \ldots,\ i_nj_n}) \\ &= \text{direct additive effect of allele } A_{i_1}, \end{aligned}$$

$$\begin{aligned} {}_d\delta_{i_1j_1} &= {}_{i_1j_1}d_{..,\_\_,..} - {}_d\alpha_{i_1} - {}_d\alpha_{j_1} \\ &= \text{direct dominance effect of } A_{i_1}A_{j_1}, \end{aligned}$$

$$\begin{aligned} {}_a\alpha_{i_2} &= {}_{..}d_{i_2.,\_\_,..} = \sum p_{i_1}p_{j_1}p_{j_2} \ldots p_{i_n}p_{j_n}({}_{i_1j_1}d_{i_2j_2,\ \ldots,\ i_nj_n}) \\ &= \text{associate additive effect of allele } A_{i_2}, \end{aligned}$$

$$\begin{aligned} {}_a\delta_{i_2j_2} &= {}_{..}d_{i_2j_2,\_\_,..} - {}_a\alpha_{i_2} - {}_a\alpha_{j_2} \\ &= \text{associate dominance effect of } A_{i_2}A_{j_2}, \text{ and} \end{aligned}$$

$$\begin{aligned} {}_{da}(\alpha\alpha)_{i_1i_2} &= {}_{i_1}.d_{i_2.,\_\_,..} - {}_d\alpha_{i_1} - {}_a\alpha_{i_2} \\ &= \text{additive} \times \text{additive interaction between the alleles } A_{i_1} \text{ and } A_{i_2}. \end{aligned}$$

The total genotypic variance may be partitioned as follows:

$$\sigma_G^2 = {}_{dd}\sigma_A^2 + {}_{dd}\sigma_D^2 + (n-1)({}_{aa}\sigma_A^2 + {}_{da}\sigma_{AA}^2 + {}_{da}\sigma_{DA}^2) + (n-1)({}_{aa}\sigma_D^2 + {}_{da}\sigma_{AD}^2 + {}_{da}\sigma_{DD}^2) + \ldots,$$

where

$$\begin{aligned} \sigma_G^2 &= \sum p_{i_1}p_{j_1} \ldots p_{i_n}p_{j_n}({}_{i_1j_1}d_{i_2j_2,\ \ldots,\ i_nj_n})^2, \\ {}_{dd}\sigma_A^2 &= 2\sum p_{i_1}({}_d\alpha_{i_1})^2, \\ {}_{dd}\sigma_D^2 &= \sum p_{i_1}p_{j_1}({}_d\delta_{i_1j_1})^2, \\ {}_{aa}\sigma_A^2 &= 2\sum p_{i_2}({}_a\alpha_{i_2})^2, \\ {}_{aa}\sigma_D^2 &= \sum p_{i_2}p_{j_2}({}_a\delta_{i_2j_2})^2, \text{ etc.} \end{aligned}$$

For prediction purposes the following covariance between direct and associate additive effects must be defined:

$${}_{(da)}\sigma_A = 2\sum p_{i_1}({}_d\alpha_{i_1})({}_a\alpha_{i_1}).$$

For greater elaboration of the model, see Griffing (1967).

(*b*) *Consequences of Selection Applied to Groups of Size* n_1 *and Evaluated with Regard to Groups of Size* n_2

As in the previous paper both individual and group selection will be considered. The arguments presented below are similar to those given in the first paper and therefore only the results of these arguments are given.

(i) *Individual Selection*

The selection value of a particular genotype, $A_{i_1}A_{j_1}$, when averaged over all groups of size n_1 is:

$$w_{i_1j_1} = 1+{}_{n_1}(\bar{\imath}/\sigma)_{\text{ind.}}[{}_{i_1j_1}d^{(n_1)}_{\cdots,\_\_,\cdots}],$$

where $\bar{\imath}$ = standardized selection differential, σ = phenotypic standard deviation, and the subscript "ind." indicates that $\bar{\imath}$ and σ relate to individual observations. The superscript (n_1) indicates that the genotypic value refers to groups of size n_1.

Following selection, the change in frequency of the allele A_{i_1} is

$$\Delta({}_{n_1}p_{i_1}) = {}_{n_1}(\bar{\imath}/\sigma)_{\text{ind.}}({}_{p_{i_1}})({}_{dn_1}\alpha_{i_1}).$$

If the selection procedure is evaluated with regard to a population of groups of size n_2, the change in mean is approximately given as follows:

$$\begin{aligned} {}_{n_1}(\Delta\mu)_{n_2} &= 2[{}_{n_1}(\bar{\imath}/\sigma)_{\text{ind.}}]\{E({}_{dn_1}\alpha_{i_1})[{}_{dn_2}\alpha_{i_1}+(n_2-1)({}_{an_2}\alpha_{i_1})]\} \\ &= {}_{n_1}(\bar{\imath}/\sigma)_{\text{ind.}}[{}_{(dn_1dn_2)}\sigma_A+(n_2-1){}_{(dn_1an_2)}\sigma_A], \end{aligned}$$

where

$${}_{(dn_1dn_2)}\sigma_A = 2E({}_{dn_1}\alpha_{i_1})({}_{dn_2}\alpha_{i_1}) = 2\sum p_{i_1}({}_{dn_1}\alpha_{i_1})({}_{dn_2}\alpha_{i_1})$$

and

$${}_{(dn_1an_2)}\sigma_A = 2E({}_{dn_1}\alpha_{i_1})({}_{an_2}\alpha_{i_1}) = 2\sum p_{i_1}({}_{dn_1}\alpha_{i_1})({}_{an_2}\alpha_{i_1}).$$

Since gene effects defined for groups of one size may be different from similarly defined effects for groups of a different size, ${}_{dn_1}\alpha_{i_1}$ need not be the same magnitude, nor even have the same sign, as ${}_{dn_2}\alpha_{i_1}$. Hence, it follows that the change in mean, ${}_{n_1}(\Delta\mu)_{n_2}$, can be negative because either one or both of the covariances $[{}_{(dn_1dn_2)}\sigma_A$ or ${}_{(dn_1an_2)}\sigma_A]$ can be negative.

(ii) *Group Selection*

With group selection, the entire group is accepted or rejected on the basis of its group mean. With such a selection procedure operating among groups of size n_1, it can be shown that the change in the frequency of the allele A_{i_1} is

$$\Delta({}_{n_1}p_{i_1}) = {}_{n_1}(\bar{\imath}/\sigma)_{\text{gr.}}({}_{p_{i_1}})\{(1/n_1)[{}_{dn_1}\alpha_{i_1}+(n_1-1){}_{an_1}\alpha_{i_1}]\},$$

where the subscript "gr." indicates that $\bar{\imath}$ and σ are group parameters.

When group selection is evaluated in terms of groups of size n_2, the change in the population mean is

$$_{n_1}(\Delta\mu)_{n_2} = (2/n_1)[_{n_1}(i/\sigma)_{\text{gr.}}]\{E[_{dn_1}\alpha_{i_1}+(n_1-1)_{an_1}\alpha_{i_1}][_{dn_2}\alpha_{i_1}+(n_2-1)_{an_2}\alpha_{i_1}]\},$$

which may be recast in terms of covariance components as

$$\begin{aligned}_{n_1}(\Delta\mu)_{n_2} = (1/n_1)[_{n_1}(i/\sigma)_{\text{gr.}}]\{&(_{dn_1dn_2})\sigma_A+(n_2-1)_{(dn_1an_2)}\sigma_A+(n_1-1)_{(an_1dn_2)}\sigma_A\\&+(n_1-1)(n_2-1)_{(an_1an_2)}\sigma_A\}.\end{aligned}$$

It is obvious that when $n_1 = n_2$ the change is a function of a sum of squares and therefore is non-negative. However, if $n_1 \neq n_2$ this no longer holds and a negative change is possible.

(c) *Selection Operating with Respect to Sets of* k *Groups each of Size* n

(i) *Population Specification*

Consider a population of groups of size n, such that genotypes within groups interact in any manner whatsoever but genotypes in different groups do not interact. The further extension now considered is to generate a population of sets such that each set contains k groups at random.

The basic unit, then, is a set having $N = (kn)$ genotypes. Such a set can be characterized as follows:

$$S = [(A_{i_1}A_{j_1}, \ldots, A_{i_n}A_{j_n})_1, (A_{k_1}A_{l_1}, \ldots, A_{k_n}A_{l_n})_2, \ldots, (A_{r_1}A_{s_1}, \ldots, A_{r_n}A_{s_n})_k].$$

Since elements in different groups are independent, the appropriate model is that previously described for groups of order n.

(ii) *Consequences of Individual Selection*

With individual selection the results are the same as those with groups of size n. Thus the change in gene frequency is

$$\Delta p_{i_1} = (i/\sigma)_{\text{ind.}}(p_{i_1})(_d\alpha_{i_1}),$$

and the change in the mean, as measured in a population of groups of size n, is

$$\Delta\mu = (i/\sigma)_{\text{ind.}}\{_{dd}\sigma_A^2+(n-1)_{(da)}\sigma_A\}.$$

This result indicates that higher-order organization of individuals into sets does not change the consequences of individual selection, if interaction between genotypes in different groups does not exist.

(iii) *Consequences of Selection Based on Sets of Groups*

When selection is based on the acceptance or rejection of entire sets, the change in gene frequency can be shown to be

$$\Delta p_{i_1} = (\bar{i}/\sigma)_N(p_{i_1})(1/N)[{}_d\alpha_{i_1}+(n-1)_a\alpha_{i_1}],$$

where the subscript N indicates that $\bar{i}$ and σ relate to the population of sets each with $N = kn$ elements.

The change in mean of the population of groups of size n is approximately given as follows:

$$(\Delta\mu)_N = (\bar{i}/\sigma)_N(1/N)[{}_{dd}\sigma_A^2+2(n-1)_{(da)}\sigma_A+(n-1)^2{}_{aa}\sigma_A^2].$$

(iv) *Comparison of Selection Based on Sets with that Based on Groups*

Recall that when selection is based on groups of size n, the change in the mean of the population of similarly sized groups is

$$(\Delta\mu)_n = (\bar{i}/\sigma)_n(1/n)[{}_{dd}\sigma_A^2+2(n-1)_{(da)}\sigma_A+(n-1)^2{}_{aa}\sigma_A^2].$$

Hence, the efficiency of set selection relative to group selection is

$$\frac{(\Delta\mu)_N}{(\Delta\mu)_n} = \frac{(\bar{i}/\sigma)_N(1/N)}{(\bar{i}/\sigma)_n(1/n)},$$

which, if $\bar{i}_N = \bar{i}_n$, becomes $(1/k)(\sigma_n/\sigma_N)$. Then, since

$$\sigma_n = \{(1/n)[\sigma_P^2+(n-1)({}_P\sigma_P)]\}^{\frac{1}{2}},$$

and

$$\sigma_N = \{(1/nk)[\sigma_P^2+(n-1)({}_P\sigma_P)]\}^{\frac{1}{2}},$$

where σ_P^2 = phenotypic variance for elements within groups, and ${}_P\sigma_P$ = phenotypic covariance between elements within groups, the efficiency of selecting among sets relative to selecting among groups reduces to a remarkably simple formula, i.e.

$$(\Delta\mu)_N/(\Delta\mu)_n = k^{\frac{1}{2}}/k.$$

This ratio is independent of the value of n. The following tabulation gives various values of this ratio for corresponding values of k:

k	1	2	4	9	16	25	100
$(\Delta\mu)_N/(\Delta\mu)_n$	1	0·71	0·50	0·33	0·25	0·20	0·10

It is clear that selection among groups (i.e. $k = 1$) is the most efficient procedure. Selecting on the basis of sets of groups can quickly reduce the efficiency of selection.

Using a similar argument, it can be demonstrated that in the absence of genotypic interaction between elements within groups, selection among groups is inefficient relative to individual selection. To show this, start with the change in the population mean due to group selection, i.e.

$$(\Delta\mu)_n = (\bar{i}/\sigma)_n(1/n)\{{}_{dd}\sigma_A^2+2(n-1)_{(da)}\sigma_A+(n-1)^2{}_{aa}\sigma_A^2\},$$

which may be recast as

$$\frac{(\bar{i}_n)(1/n)\{{}_{dd}\sigma_A^2+2(n-1)_{(da)}\sigma_A+(n-1)^2{}_{aa}\sigma_A^2\}}{\{(1/n)[\sigma_P^2+(n-1)({}_P\sigma_P)]\}^{\frac{1}{2}}}.$$

If individuals within groups are independent, this becomes

$$(\Delta\mu)_n = \bar{i}_n(n^{\frac{1}{2}}/n)({}_{dd}\sigma_A^2/\sigma_P).$$

Under the assumption of no genotypic interference, individual selection results in the following change in mean:

$$(\Delta\mu)_1 = \bar{i}_1({}_{dd}\sigma_A^2/\sigma_P).$$

If $\bar{i}_n = \bar{i}_1$, the relative efficiency of group selection in comparison with individual selection is

$$(\Delta\mu)_n/(\Delta\mu)_1 = n^{\frac{1}{2}}/n.$$

This illustrates that although group selection ensures non-negative changes in the population mean, it can (under certain circumstances) be an inefficient form of selection.

III. Discussion

Extension of the biological model to accommodate genotypic interaction necessitates the generation of populations of groups. This permits the additional complexity that selection may operate on groups of one size and be evaluated with regard to groups of another size. In this situation the present analyses indicate that even with group selection there can be no assurance that positive selection will invariably lead to a non-negative change in the population mean. Hence, if invariably a non-negative change is desired, selection should be carried out in populations having a group size similar to that in which selection is to be evaluated.

As a practical illustration consider a pasture-breeding problem. From the analyses presented in this study it is clear that positive selection in spaced plants can lead to a negative response as measured under sward conditions. This is especially true of those traits that are subject to competitional stress. Perhaps this is one reason why pasture breeding in the past has not resulted in greater fulfilment of objectives.

Limiting discussion to the case of a population with a fixed group size, the first paper of the series demonstrated that if direct and associate effects were negatively correlated, positive individual selection could lead to a negative response.

However, it was shown that this dilemma could be resolved by transferring the basis of selection from that of the individual to that of the group. In the present study, it is demonstrated that in some instances the cost to ensure a non-negative response can be high. For those traits which exhibit no genotypic interaction, the efficiency of group relative to individual selection is given by $n^{\frac{1}{2}}/n$. Hence, the optimum procedure is to derive a selection theory based on an index which combines individual and group selection in such a way as to invariably yield the maximum possible non-negative change in the population mean. This problem will be discussed in future contributions to this series.