EFFECTS OF FINITE SIZE ON SELECTION ADVANCE IN SIMULATED GENETIC POPULATIONS*

By J. L. GILL†

[*Manuscript received October* 26, 1964]

Summary

The design and analysis of an investigation of selection in simulated genetic populations is discussed. The long-range objectives and specification of parameters and models are outlined, and certain mechanics of simulation of genetic systems are included, some of which may be of general application to Monte Carlo investigations.

The results from selection in small simulated populations under nine different models of gene action are discussed with respect to the effects of population size on inbreeding depression and the random drift of gene frequency.

I. INTRODUCTION

A major deficiency of quantitative genetics is the void between the mathematician working with simple genetic models and the experimenter working with organisms of extreme genetic complexity. A mathematical description of the complex genetic situations results in equations too cumbersome for solution, and simplifying assumptions usually lead to large departures from reality. In recent years, some numerical methods which circumvent some of those difficulties have been introduced into the field of quantitative genetics. The technique which allows one to simulate the processes of genetics through the use of repetitive sequences involving random numbers generated by a high-speed computer has been termed the Monte Carlo method. This approach enables one to study the joint effects of dominance, epistasis, linkage, and other variables upon the progress of finite genetic populations under selection.

Fraser (1957*a*, 1957*b*) introduced the first techniques for the simulation of genetic systems by digital computers and studied the effects of linkage on the rate of advance under selection. His colleague Barker (1958*a*, 1958*b*) investigated selection between alleles at autosomal and sex-linked loci, showing that it is possible to simulate the operations of natural selection. Fraser (1960*a*, 1960*b*, 1960*c*) also simulated selection against phenotypic extremes and used non-linear transformations to study epistasis, although he admitted that it is more generally valid to base any analysis of epistasis on the interactions of individual genes.

Martin and Cockerham (1960) and Bohidar (1960) conducted investigations that involved Monte Carlo applications to quantitative genetic studies of selection

* Journal Paper No. J-4804, Iowa Agricultural and Home Economics Experiment Station, Ames. Project No. 1508, supported by National Science Foundation Grant G-19218.

† Statistical Laboratory, Iowa State University, Ames; present address: Department of Dairy, Michigan State University, East Lansing, Michigan.

without epistasis. In both studies, tight linkage slowed progress from selection or resulted in plateaus of genetic merit.

Baker (1961) presented results of a Monte Carlo study directed towards evaluating the premise that the average genotypic value of the heterozygote at a specific locus can be enhanced by the effects of linked loci under specified conditions. The results from populations with dominance, but no epistasis, suggest that linkage aids in the persistency of polymorphism but does not hinder the progress of the genotypic mean, except in very small populations. Lewontin and Dunn (1960) and Crosby (1960) also have used computers to simulate genetic effects.

Despite differences in specification of simulation procedure and in choice of goals and parameters, these studies indicate that the basic simplicity and unusual applicability of the Monte Carlo technique to quantitative genetics make it a valuable tool for the study of the forces involved in inheritance.

II. Objectives

This study is part of an investigation undertaken as a feasible approach to increasing the understanding of the parameters involved in genetic selection. The broad objective is to determine the effects of population size, degree of truncation selection, environmental variation, linkage, and mode of gene action upon the progress of genetic populations under selection. Obviously, a comprehensive study of these effects should include an attack on the interactions among them and comparisons of various approaches to simulation of realistic situations. This involves a series of coordinated investigations of which the earlier efforts, involving the general effects of a wide variety of parameters, will be used as a guide for the concentration of subsequent efforts in areas deemed potentially most fruitful.

The specific results of initial interest are the genotypic and phenotypic means and variances of the selected and unselected populations in each generation, the amount of fixation and frequencies of unfixed genes, changes in components of variance, achieved selection differentials, and the heritability realized from selection. From these data the mode of operation of each parameter and the general force exerted on the population under selection can be investigated.

In this paper the general simulation procedures will be outlined, and some results will be discussed with respect to population size.

III. Specification of Parameters and Models

Unisexual diploid individuals were simulated, their quantitative characteristics assumed to be expressed in both sexes. A metric characteristic was determined by the genes at 40 loci equally spaced over eight chromosomes, with two alleles per locus and equal genetic effect at all loci. Should one choose to make the magnitude of effects of some genes larger than others, the most likely consequence would seem to be more rapid transition toward fixation of genes with large effects without affecting total progress from selection. The ratio of gene effect to total phenotypic variation has been suggested as an important parameter. However, such a quantity could change considerably under long-term selection, and, therefore, may not be suitable as

a fixed parameter. Equal numbers of parents of each sex, selected by upper truncation of phenotypes, were mated at random by sampling with replacement. The populations were assumed to be free from the effects of mutation and natural selection over a period of 30 non-overlapping generations.

The effective population sizes specified (8, 12, 16, or 32 parents) should produce, in the absence of other forces, reductions in the panmictic index of 6·25%, 4·17%, 3·12%, and 1·56% respectively per generation because of the finite size of the population.

The selection intensities specified ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{6}$, or $\frac{1}{8}$), where combined with the specified numbers of parents, determine progeny populations ranging from 16 to 256 in number, and correspond to selected population means which are expected to be 0·8, 1·27, 1·5, and 1·65 standard deviations respectively above the mean of an unselected population that is conceptually infinite and normally distributed. The expected selection differentials are somewhat smaller, but of similar proportions for small populations, where lack of extremes produces some non-normality. Selection intensity was constant for a given run, or parameter set, from one generation to the next for 30 generations or until fixation occurred at all loci.

The amount of simulated environmental variation was specified at four levels related to the expected genotypic variance in the initial generation of progeny (0, $\sigma_G^2/3$, σ_G^2, and $3\sigma_G^2$). If one desires to discuss heritability in terms of 40 loci, then these levels correspond to initial heritabilities, in the broad sense, of 1·0, 0·75, 0·50 and 0·25 respectively.

To simulate linkage, recombination values of 0·005, 0·05, 0·2, and 0·5 were applied to the adjacent loci on each chromosome, with the probability of crossover being uniform for all adjacent pairs of loci on the same chromosome for a given run. Crossover interference was not simulated.

Epistasis was restricted to interactions of sequential pairs of loci. Fisher (1918) expressed the opinion that interactions more complex than the dual type were of little importance. However, Lush (1948) pointed out that the number of possible interactions is so large that, if only a few of the more complex ones are real, they could furnish a large amount of epistatic variance. The additional restriction of epistasis to interactions of sequential pairs of adjacent loci may be more serious because such a procedure accounts for only a small fraction of all possible interacting pairs. If n loci exist, $n(n-1)/2$ interactions between two loci are possible. Simulation of $n/2$ sequential interactions accounts for only $1/(n-1)$ of the possible total of dual interactions.

The models of gene action that were studied are listed in Table 1. Genotypic values for a pair of loci are given, along with the expected genotypic means and variances of the offspring created from an initial parent population that is completely heterozygous, but has random association of coupling and repulsion linkage phases. The effects of different fixed amounts of initial linkage disequilibrium may be of interest but are not considered in this study.

The additive, complete dominance, and overdominance models are standard non-epistatic models, although the overdominance model assumed is one specific

case in which the two homozygotes have equal values. The remaining six models contain epistasis in some form. Partitions of the total genotypic variance, under

TABLE 1

GENETIC MODELS SIMULATED

Model	Genotypic Values				Expected Initial Values for 40 Loci	
		AA	*Aa*	*aa*	Mean	Variance
Additive	*BB:*	12	11	10	200	20
	Bb:	11	10	9		
	bb:	10	9	8		
Complete dominance	*BB:*	11	11	9	200	30
	Bb:	11	11	9		
	bb:	9	9	7		
Overdominance	*BB:*	8	10	8	200	40
	Bb:	10	12	10		
	bb:	8	10	8		
Optimum number	*BB:*	7	10	11	200	30
	Bb:	10	11	10		
	bb:	11	10	7		
Duplicate factors	*BB:*	10	10	10	195	18·75
	Bb:	10	10	10		
	bb:	10	10	6		
Complementary factors	*BB:*	11	11	9	202·5	20
	Bb:	11	11	9		
	bb:	9	9	9		
Additive-by-additive conditional epistasis	*BB:*	12	10	8	200	20
	Bb:	10	10	10		
	bb:	8	10	12		
Additive-by-dominance conditional epistasis	*BB:*	12	9	10	200	20
	Bb:	9	10	11		
	bb:	10	11	8		
Dominance-by-dominance conditional epistasis	*BB:*	9	11	9	200	20
	Bb:	11	9	11		
	bb:	9	11	9		

random-mating conditions, for a generalized gene model with epistasis but no linkage have been provided by Kempthorne (1954) and Cockerham (1954), although Cockerham's procedure is limited to consideration of only two alleles per locus.

The relations of the components and the means to gene frequency for each of the six epistatic models are given by Figures 1–6.

The optimum gene number model, first proposed by Wright (1935), is based on the equation

$$\text{Yield} = C - (\text{number of plus alleles} - K)^2$$

for a pair of interacting loci. In this study C was set at 11 and $K = 2$ for intermediate preferred. With the duplicate factor model, substitution of either gene can produce the outward phenotypic change, but does not do so if the other has already produced the effect. With the complementary factor model, neither gene can produce an effect unless the other is present.

Three conditional epistatic models were constructed by restricting the generalized gene model of Kempthorne (1957, p. 416) to single components of epistatic variance at a particular gene frequency, in this case, one-half.

IV. Design and Analysis

Populations were simulated for each of 16 runs, or parameter sets, associated with each of the nine models of gene action. The content of each of the parameter sets was derived from the orthogonal array of a $\frac{1}{16}$ fractional replication of a 4^4 factorial design. The four factors are environmental variation, population size, selection intensity, and linkage. Estimates of the main effects are valid only if the aliases, all of which are interactions, are negligible. Comparisons among the main effects are orthogonal, however, because each level of a factor occurs the same number of times (once in this plan) with every level of the other factors in some parameter set of the design. The two-factor interactions are not estimable. The main effect of linkage is not confounded with any two-factor interactions. The existence of interactions of selection intensity with environmental variation and with population size has been inferred from selection experiments. Thus, these interactions may be of particular importance in relation to inferences about the main effects of population size and environmental variation, respectively. No other interactions of known importance are confounded with main effects.

Each gene model was studied with a separate fractional plan. Therefore, inferences concerning differences in main effects among models rest on a firmer basis than do those concerning differences within a particular model.

The analysis of results of each population parameter set for each model was performed after averaging the results over five consecutive generations. The grouping seemed desirable in order to eliminate some of the random sampling from generation to generation and to reduce the unwieldy amount of data. It is true that the consecutive statistical tests are correlated, but the exact relationship is not clear because of the varying amount of fixation that occurs. The differences among levels of any factor could have been tested by using the mean square for pooled interactions for an error term. However, by repeating the basic 16 parameter sets using different random starts in the computer, the tests were made with more precision, and an additional test could be made of the significance of the mean square for pooled

interactions. The analysis of variance is straightforward and allows for equal precision in the tests for all factors.

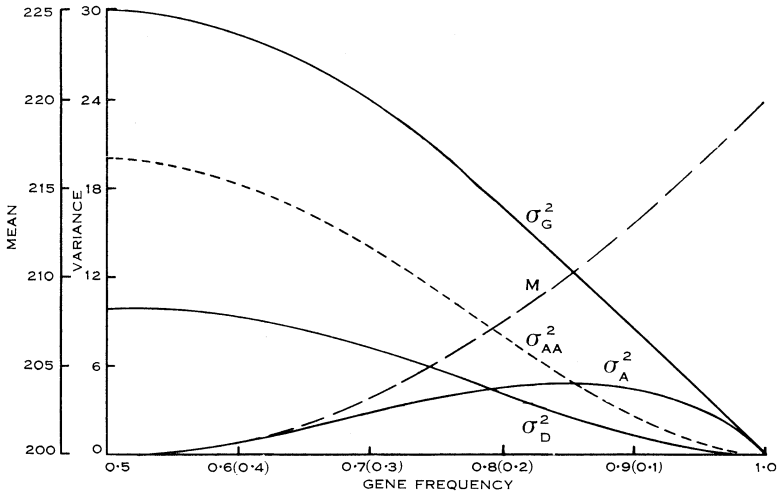

Fig. 1.—Changes in the mean (M), total genotypic variance (σ^2_G), and components of variance under the optimum number gene model, when changes in the frequencies of two interacting genes are of equal magnitude but opposite in direction.

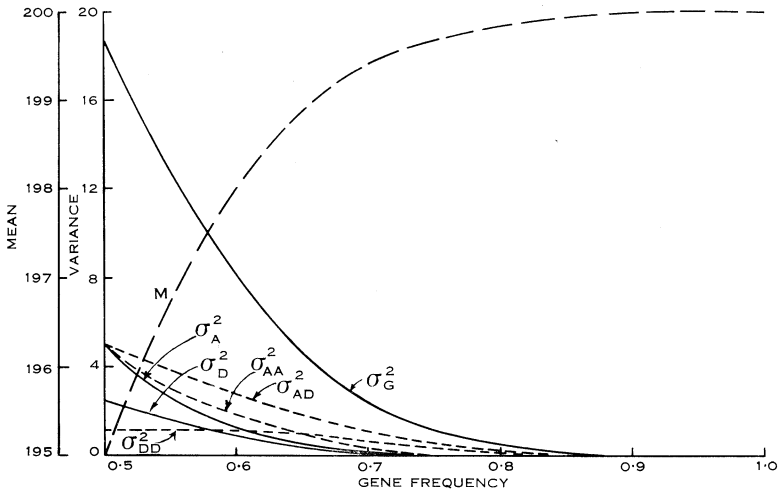

Fig. 2.—Changes in the mean (M), total genotypic variance (σ^2_G), and components of variance under the duplicate factor model, when the change in gene frequency is the same at all loci.

V. Mechanics of Simulation

In considering the actual computer programme, only those points of methodology and logic that seem potentially useful to others will be discussed.

Although there are many different types of Monte Carlo analyses, they all include as a common feature the use of sets of pseudo-random numbers. The multi-

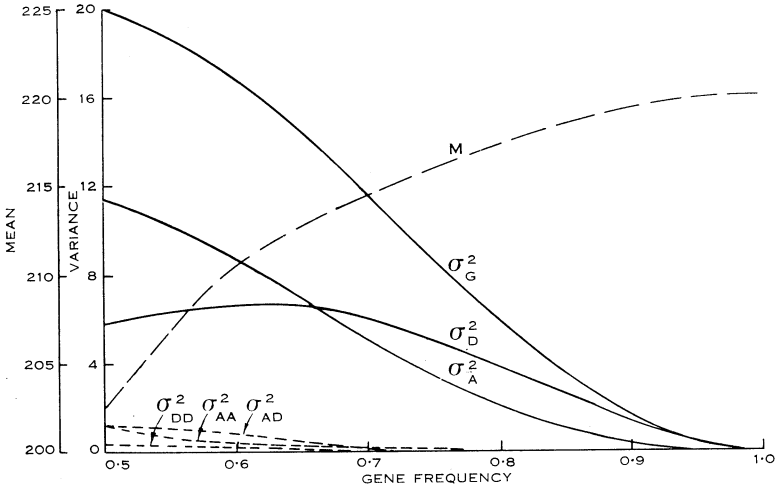

Fig. 3.—Changes in the mean (M), total genotypic variance (σ^2_G), and components of variance under the complementary factor model, when the change in gene frequency is the same at all loci.

plicative congruential method for generating uniformly distributed pseudo-random numbers on a digital computer was proposed by Lehmer in 1951. The procedure

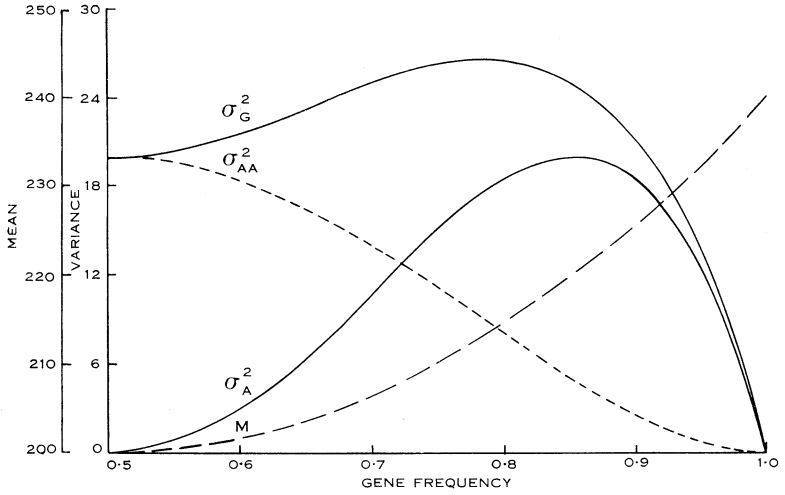

Fig. 4.—Changes in the mean (M), total genotypic variance (σ^2_G), and components of variance under the additive-by-additive conditional epistatic model, when the change in gene frequency is the same at all loci.

was modified by Rotenberg (1960) and Greenberger (1961). Results of tests of the validity of this method and comments on the choice of parameters are given by Tausskey and Todd (1956), Rotenberg (1960), and Peach (1961).

In the simulation of environmental effects, random normal deviates must be generated, or tabulated values stored in the computer. Efficiency of time and

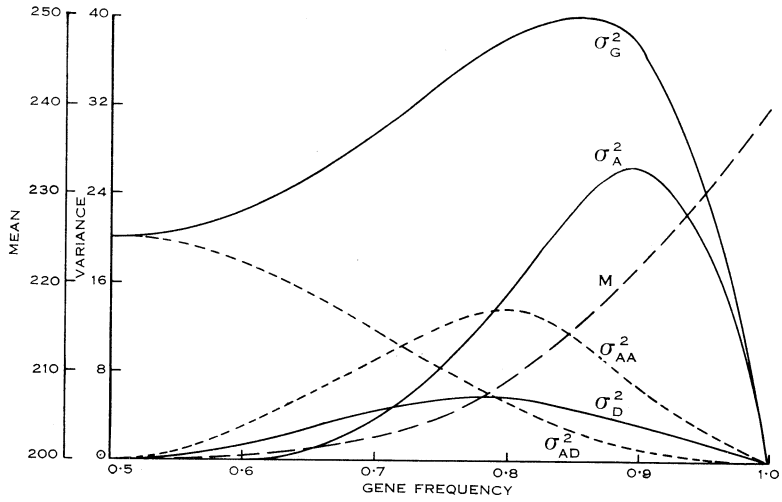

Fig. 5.—Changes in the mean (M), total genotypic variance (σ^2_G), and components of variance under the additive-by-dominance conditional epistatic model, when the change in gene frequency is the same at all loci.

space may be obtained by generating the deviates with a relatively simple process based on the central limit theorem. In this study seven uniformly distributed

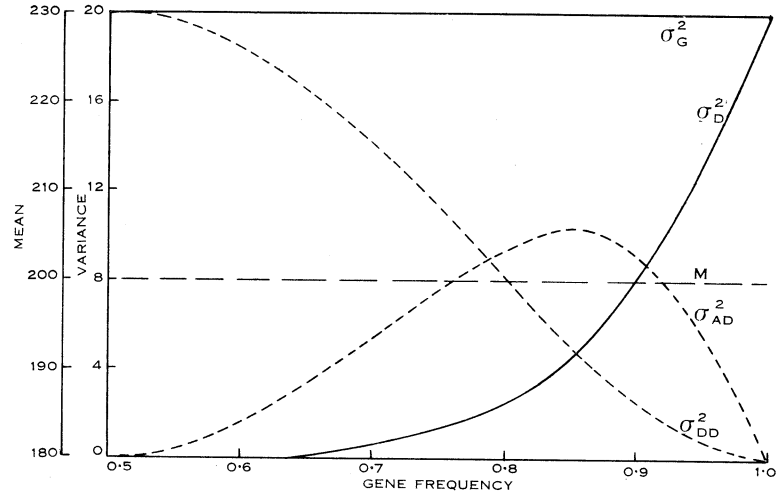

Fig. 6.—Changes in the mean (M), total genotypic variance (σ^2_G), and components of variance under the dominance-by-dominance conditional epistatic model, when the frequency of a certain gene changes and that of an interacting gene remains constant at 0·5.

random numbers in the range $-1<r_i<+1$ were generated and summed; the sums were coded to fit the normal distribution with mean zero and unit variance by

multiplying by 3/7. Those numbers then are in the range $-7<e_j<+7$. Muller (1959) has given the advantages and disadvantages of this method for generating random normal deviates and discussed several alternative procedures.

In the process of generating the initial parents, random numbers are used to choose 1- or 0-alleles along the first chromosome of each pair with equal probability at each locus. By taking the complement of this array of 40 alleles, the corresponding alleles of the second chromosome of each pair are generated so that they are alternate to those of the first chromosome at every locus, producing the desired completely heterozygous individual. Other initial gene frequencies or genetic structures of the population also may be of interest but were not considered here.

After the initial generation, parents are chosen randomly from the selected group of each sex, and the processes of segregation, recombination, and fertilization are simulated by a random mask method which produces a gamete from each parent. This method was first developed by Schweppe and Bohidar at Iowa State University, and was used in a study by Bohidar (1960). The method is faster than the random transform of a vector of genetic recombination frequencies, which Fraser (1957*a*) used for a small number of loci, and is very much faster than the random walk method which he proposed for linkage of large numbers of loci (Fraser 1960*a*).

Techniques involving logical algebra, which Fraser (1957*a*) has described, may be used to identify genotypes to evaluate them numerically. In the evaluation of phenotypes, random normal deviates which simulate environmental values are added to the numerical genotypic values to create phenotypic values.

The selection process involves upper truncation of a given fraction $b = n/N$ of the population of ordered phenotypic values. The mechanics of the procedure are complicated because the values being ordered cannot be used directly in the reproduction simulation in the next generation. The "Bucharest Sort" described by Evans and Perry (1960) is an efficient method if the identity of each genotype can be retained in the same memory unit with the phenotypic value.

The total computer time involved in 30 generations of selection, reproduction, and evaluation ranged from approximately 3 min for populations of 16 from which eight were selected to almost 25 min for populations of 256 from which 32 were selected; i.e. the generation interval varied from 6 to 50 sec.

VI. Effects of Population Size

A certain amount of non-random mating occurs in populations of restricted size, resulting in the accumulation of inbreeding effects despite the lack of purposeful inbreeding.

(*a*) *Random Drift of Gene Frequency*

Wright (1931) has shown that heterozygosity decreases by a proportion which is approximately $1/2N$ each generation, where N is the size of the breeding population. The amount of heterozygosity is proportional to $P = (1-F)$ for inbred populations derived from random-mating populations, where F is Wright's coefficient of inbreeding. Wright (1951) has called F the fixation index and P the panmictic index. For

the case of random mating with effective population size equal to N, it may be shown that

$$F_t = 1/2N + (1-1/2N)F_{t-1},$$

for generation t. In terms of the panmictic index,

$$P_t = (1-1/2N)P_{t-1}.$$

After t generations, the panmictic index would become, approximately,

$$P_t = P_0 e^{-t/2N},$$

as given by Malecot (1948). For P to decrease to a fraction, x, of its value in the original generation, the number of generations required is the solution of $e^{-t/2N} = x$, or $t = -2N \log_e x$.

Under conditions of the dominance-by-dominance (D×D) conditional epistatic model used in this study, selection does not contribute to change in heterozygosity or fixation because of the lack of variation from additive effects or their interactions (Fig. 6). For example, the mean proportions of fixed loci at generation 25 were 53, 48, 49, and 50% for eight populations under each of four respective selection intensities, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{6}$, and $\frac{1}{8}$. Consequently, random genetic drift is primarily responsible for differences in the fixation of loci. Linkage might be expected to modify the rate, but that did not happen over an extended period of time. The mean proportions of fixed loci at generation 25 were 49, 51, 49, and 51% for eight populations with each of four respective recombination frequencies, 0·005, 0·05, 0·2, and 0·5.

The expected remaining heterozygosis may be calculated for $x = e^{-t/2N}$. In the simulated populations used in this study, $P_1 = 1$; i.e. the offspring in generation 1 (rather than generation 0) have the required Hardy–Weinberg distribution at each of the 40 loci. Therefore, expected values calculated for generation t correspond to observed values for generation $t+1$. Thus, if one takes $t = 24$ and $n = 8$, 12, 16, and 32, the respective values for x are 22, 37, 47, and 69%. The observed mean proportions of loci still segregating at generation 25 for eight populations of each size were 27, 46, 50, and 77% respectively. The agreement is fairly reasonable, at least from a statistical viewpoint, if allowance is made for the difference between fractional losses of heterozygotes and irrevocable loss of an allele at the same locus.

Robertson (1960) has shown that the "half-life" of gene frequency, i.e. the time required to move gene frequency halfway to the limit, is $1 \cdot 4N$ generations with weak selection for the additive case, or for dominance when the initial gene frequency is one-half, where N represents the effective population size. For stronger selection, he has suggested that the half-life decreases continuously as Ns increased, where s is the selection pressure on a particular allele, and that one may conclude that all desirable alleles have been fixed if the half-life is reached well before the range of N to $2N$ generations.

Results from simulated populations with complete dominance and initial gene frequency of 0·5 indicate that the half-life does decrease as Ns increases. For $N = 8$, 12, 16, and 32 the half-life (in generations) was $1 \cdot 4N$, $1 \cdot 0N$, $0 \cdot 8N$, and $0 \cdot 3N$ respectively, with Ns increasing from 2·7 to 10·9 over that range. No undesirable alleles

were fixed when the half-life was reached in $0 \cdot 3N$ generations. However, when the half-life was reached in $0 \cdot 8N$ generations, 15% of the alleles that were fixed were undesirable. Results were similar when complementary gene action was simulated, except that the half-life was a bit longer at comparable values of Ns.

For populations of size 32 with complete dominance, no fixation of the recessive alleles occurred in any population at any level of selection over the entire 30 generations. However, the mean fixation of dominant alleles for eight such populations was 55% at generation 25. This compares with mean fixation of 64% of the dominant alleles and 10% of the recessive alleles in populations of size 16.

With overdominance and equilibrium gene frequency of $0 \cdot 5$, fixation was equally divided between alleles of either type as it would be without selection under any model. Since mass selection and random drift are antithetical forces when overdominance exists, less fixation occurred for all sizes of population than that which occurred under other models of gene action. At generation 25 the mean percentages of loci still segregating were 33, 57, 72, and 92 for populations of 8, 12, 16, and 32 breeding individuals, respectively. The corresponding values under the $D \times D$ model (where selection is ineffective in all populations) of 27, 46, 50, and 77% indicate that the mean level of selection utilized with overdominant populations is considerably more effective in reducing fixation in larger populations than in small ones. The size of population needed to prevent the loss of alleles will, of course, depend on the selection pressure; i.e. 30 individuals were sufficient when selection was strong ($\frac{1}{8}$) but more numbers were required for lower intensities.

The fixation results observed for populations under the complementary factor model were quite similar to those for populations with complete dominance.

Under all other models of epistatic nature (optimum number, duplicate factors, $A \times A$, and $A \times D$) the rate of fixation was rapid and relatively invariant with increase in population size from eight to 16, but dropped noticeably with an increase to 32 breeding individuals. Under models where one allele is favoured over the other, epistasis contributed significantly to loss of favoured alleles through random extinction.

Robertson (1961) indicated that inbreeding will be greater than that calculated from the actual number of parents when both the intensity of selection and the heritability of the character are high. The standard panmictic index prediction indicates that, over a period of 24 generations, populations of eight individuals should lose $2 \cdot 5$ times as much of the original heterozygosity as populations of 32. Simulated populations of those sizes were compared for the similar characteristic, rate of loss of alleles, under selection intensities of $\frac{1}{2}$ and $\frac{1}{8}$ for 25 generations. Unfortunately, because of the fractionation of the experimental design, the amounts of environmental variation and linkage differed and could not be cancelled from the comparisons, although the initial heritabilities, in the broad sense, were between $0 \cdot 5$ and $0 \cdot 75$ in the populations being compared. The ratio of alleles lost under the milder selection pressure ($\frac{1}{2}$) was nearly as predicted, but with stronger selection ($\frac{1}{8}$), the smaller populations (8) lost more than three times as many alleles as the larger ones. Comparisons involving wider differences in population size and selection

intensity with standardized environmental variation and linkage might support Robertson's hypothesis more conclusively.

(*b*) *Inbreeding Depression Due to Population Size*

The random drift of gene frequency, or inbreeding effect due to the finite size of the population, obviously is related to changes in the mean of the population. Lush (1948), for example, shows that the phenomenon known as inbreeding depression is dependent on the degree of dominance, being zero for the case of no dominance and possibly quite large for extreme overdominance.

Kojima (1961) has shown that size of population does not cause a serious difference between the gain from selection expected in an infinite population according to the standard prediction equation, selection differential $\times$ heritability, and the expected gain from his formula for a small population, except when dominance exists. He pointed out that the joint effects of the finite size of population and dominance could amount to considerable bias in the standard prediction equation; such bias largely can be accounted for by inbreeding depression. The bias is larger, of course, when overdominance is present.

Kempthorne (1957, p. 444) expressed the mean of an inbred population, μ_I, in terms of the mean of the original random-mating population, μ_R, and the average dominance deviations of homozygotes in that population, by the equation

$$\mu_I = \mu_R + FD_1 + F^2D_2 + F^3D_3 + \ldots,$$

where the subscripts of D's refer to the number of loci involved in a particular type of dominance effect and F is Wright's coefficient of inbreeding. If there are no epistatic deviations that involve only dominance (i.e. no D_2, D_3, etc.) then the mean of the inbred population is linearly related to F in spite of the existence of other epistatic deviations.

The effect of population size on genetic progress was of major importance for only four of the nine genetic models simulated; complete dominance, overdominance, complementary factors, and dominance-by-dominance ($D \times D$) conditional epistasis. It is significant (with respect to Kempthorne's expression) that those are the only models simulated which involve large proportions of variation due to dominance effects and their interactions over an extended range of gene frequency (Figs. 1–6). Figure 7 illustrates, for those models, the sensitivity to the type I statistical error, the rejection of a hypothesis that is true, with reference to tests of significance of differences among observed means of simulated populations of four sizes ($N = 8$, 12, 16, 32 parents). The purpose of this type of graphical presentation is to illustrate the probability level at which one may reject a null hypothesis of no differences among means and to show how confidence in the conclusion increases as information from additional generations becomes available. The line for each model represents the mean result from 32 populations.

Figures 8–11 illustrate the mean genetic progress of eight populations of each size with complete dominance, overdominance, complementary factor, and $D \times D$ conditional epistasis, respectively.

With complete dominance (Fig. 8) the difference between means of populations of 32 and 8 breeding individuals was the only statistically significant difference among all paired comparisons of populations of different size over the first 20 generations of selection. However, in later generations all pair differences except the one for populations of 16 versus 12 approached statistical significance at the 5% probability level, emphasizing the inbreeding depression which accumulates when dominance is present. The populations of size 16 with low selection intensity ($\frac{1}{2}$) and high environmental variation produced negative regression of mean on generation

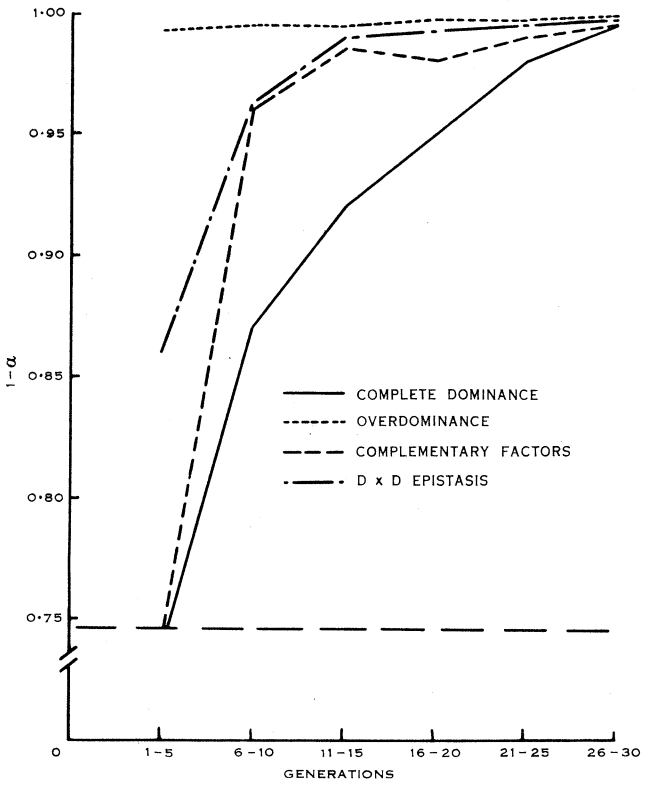

Fig. 7.—Sensitivity to type I errors in tests of hypothesis concerning the effect of population size on progress of the genetic mean.

number, and appear to be responsible for the small magnitude of the difference between means of all populations of 16 and 12 individuals, since no populations of size 12 were selected with the lower intensities under high environmental variation. This type of limitation in making comparisons is due to the fractionation of the experimental design.

With overdominance present (Fig. 9), comparisons of the means of all pairs of populations of different size revealed high statistical significance for almost every comparison over the entire period of 30 generations. If one could be certain that

overdominance was the predominant type of gene expression, inbred lines would be formed, selected, and crossed to take advantage of it, since mass selection cannot long maintain genetic superiority when favored genotypes are heterozygous. However,

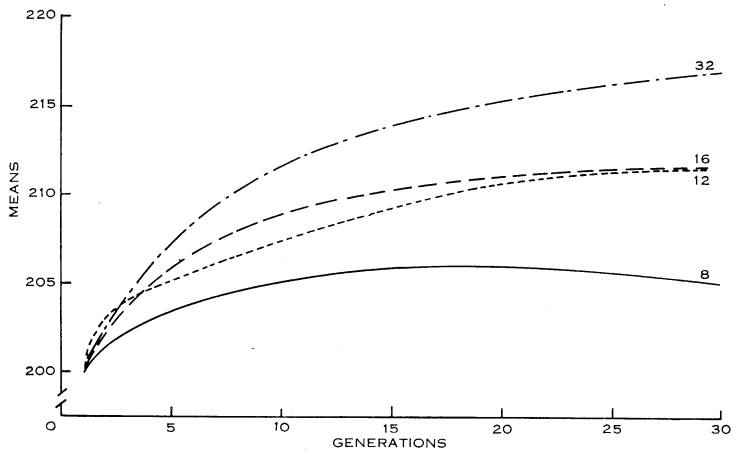

Fig. 8.—Mean genetic progress with complete dominance—by effective size of population.

if one does not know that overdominance predominates, when in fact it does, the results should be similar to those observed in the simulated populations, i.e. very strong inbreeding effects in small populations, quickly overriding the effects of

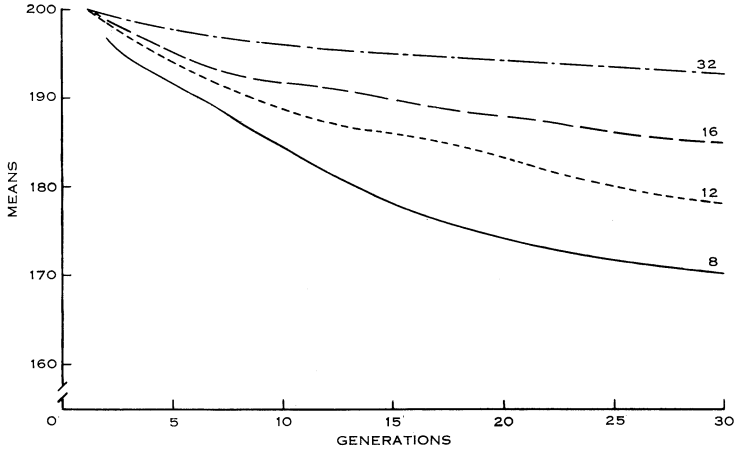

Fig. 9.—Mean genetic progress with overdominance—by effective size of population.

selection and resulting in negative regression of mean on generation number. For overdominance models with equilibrium gene frequencies other than 0·5 the results, undoubtedly, would differ somewhat in degree.

The complementary factor model is strongly suggestive of the complete dominance model (Table 1). Much of the variation is due to dominance effects and

their interactions (Fig. 3). Consequently, it is not surprising that the results of mean genetic progress by level of population size (Fig. 10) resemble those of dominance (Fig. 8).

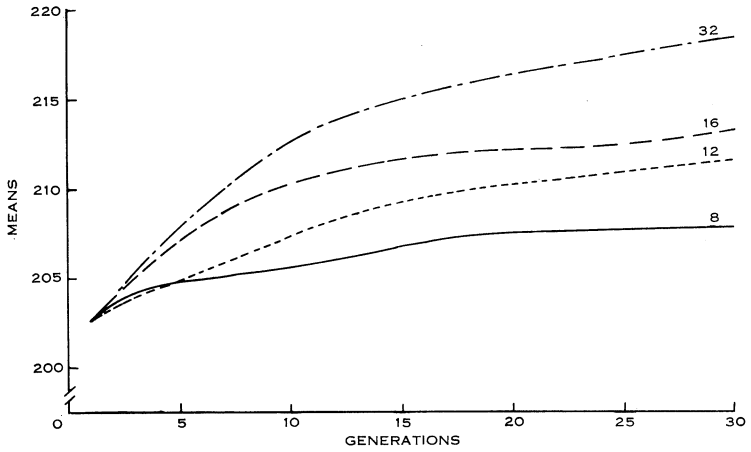

Fig. 10.—Mean genetic progress with complementary factors—by effective size of population.

The $D \times D$ conditional epistatic model involves interactions of loci exhibiting the two forms of overdominance (heterozygote superior or inferior), and is free from the effects of selection because variation due to additive effects (σ^2_A) and their inter-

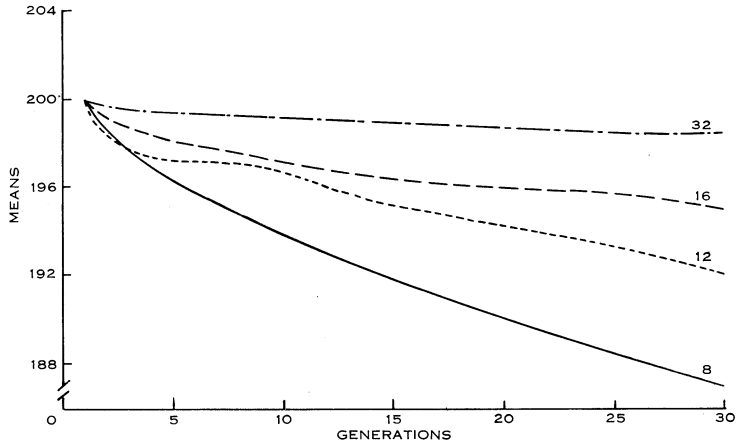

Fig. 11.—Mean genetic progress with dominance-by-dominance conditional epistasis—by effective size of population.

actions (σ^2_{AA}) do not exist in such populations at any gene frequencies (Fig. 6). Thus, the changes in genetic mean illustrated in Figure 11 are not surprising. Except for smaller magnitude, the negative regressions of mean on generation number and the differences between means of populations of different sizes are remarkably like those

illustrated for overdominance (Fig. 9). That difference in magnitude undoubtedly is conditioned by the difference in initial variation in the two models (Table 1).

With respect to size of population, none of the differences observed in populations with duplicate factors, optimum gene number model, or with additive-by-additive ($A \times A$) or additive-by-dominance ($A \times D$) conditional epistasis even remotely approached statistical significance at generally accepted probability levels.

With duplicate factors, it appears that any significant amount of artificial selection (i.e. one-half the population or fewer saved per generation) will exclude the only undesirable genotype, the double recessive homozygote, sufficiently to raise the mean consistently regardless of population size. Of the four forms which random fixation can take, *AABB*, *AAbb*, *aaBB*, or *aabb* states for interacting pairs of loci, fixation in any of the first three states contributes maximum genotypic value with duplicate gene action, in contrast to the comparable additive situation, where maximum value is achieved only in the first genotypic state. Therefore, random drift roughly has only one-third the detrimental potential, from that point of view, when duplicate factors are acting, and appears to supplement the force of selection, rather than to act antithetically.

With the optimum and $A \times A$ conditional epistatic models, selection becomes more effective at higher gene frequencies in all sizes of populations. Both models involve high proportions of σ^2_{AA} at gene frequencies near one-half, but the populations have less of that and have larger proportions of additive variance for other gene frequencies (Figs. 1 and 4). Two of the four fixation states have maximum genotypic value in either model (Table 1), so that random genetic drift may cause as many advantageous as deleterious changes, essentially cancelling its effect on the mean. Dominance variance that, theoretically, causes inbreeding depression is present to some degree under the optimum model, but does not occur in populations associated with the $A \times A$ model. With the $A \times A$ model, the existence of two homozygous peaks of genetic merit insures that gene frequencies removed from one-half will be pushed farther toward one of the limits, 0 or 1, as selection proceeds over time. In fact, genetic progress was even more rapid than that observed with additive gene action, because the force of selection predominates over random drift to the extent that populations of size eight accrue little more deleterious fixation than do those of size 32.

The $A \times D$ conditional epistatic model is a complex case, where the intra-locus effects at locus A are overdominant (heterozygote superior) when combined with *bb*, additive when combined with *Bb*, and overdominant (heterozygote inferior) when combined with *BB* (Table 1). Consequently, if selection is effective in increasing the mean frequency of dominant alleles at most loci, overdominance involving inferior heterozygotes will predominante. Then, because of the magnitude of dominance deviations, the effect of inbreeding on the mean will be almost twice as large as in the usual case of inbreeding depression with complete dominance (cf. Lush 1948, p. 260; or Falconer 1960, p. 251), but will be in the *opposite* direction. A different name is needed for this contrasting phenomenon. Perhaps one could term it "inbreeding uplift". Such a phenomenon is suggested by the results of Bennett (1960),

which indicate that selection for insecticide (DDT) resistance in *Drosophila* populations was more effective when coupled with inbreeding because of the selective advantage of homozygous types. However, if selection or random drift causes gene frequency to change in the direction of either or both of the secondary peaks of genetic merit (Table 1), considerable dominance variation is generated and inbreeding depression develops.

Although the differences in means due to population size were not statistically significant with the $A \times D$ model, the trends may portend situations which are real (Fig. 12). In the first 10–15 generations of selection, the means of the smaller populations increased considerably faster than those of the larger populations, but the

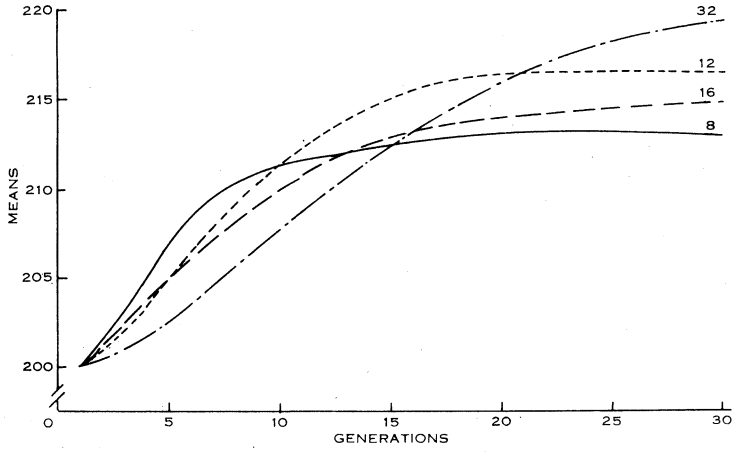

Fig. 12.—Mean genetic progress with additive-by-dominance conditional epistasis—by effective size of population.

trend was reversed in the next 15–20 generations. Initially, selection probably is effective in changing the gene frequency at many loci rather rapidly toward the homozygous primary peak of genetic merit, with assistance from inbreeding uplift. Later, with exhaustion of additive variance among those loci, selection would be applied more and more to the remaining loci, whose gene frequencies have drifted somewhat toward the secondary genetic peaks, some as far as fixation of the undesirable allelic states in smaller populations, with less of that in the larger populations. The sharp, rapid decline of the observed amount of genotypic variance, beginning about the sixth generation in the smaller simulated populations, and later, with correspondingly less severe decreases in the larger populations, is consonant with the theoretical explanation.

(c) *Conclusions*

(1) For metric characteristics influenced by complete dominance, the critical size of a simulated population, with respect to prevention of random extinction of favored alleles, was between 16 and 32 breeding individuals, when significant intensity of selection was applied over an extended period. There was some evidence against the hypothesis (Robertson 1960) that all desirable

alleles have been fixed if the "half-life" of gene frequency is reached well before the range of N to $2N$ generations.

(2) Effective populations of 30 or more individuals were needed to prevent loss of alleles by random drift when overdominance existed and $\frac{1}{8}$ or more of the total population were selected as parents.

(3) With epistatic models where one allele is favored over the other in certain combinations, interlocus interactions contributed significantly to loss of favored alleles through random extinction; under all other epistatic models the rate of fixation was rapid and relatively invariant with population size for groups of 8, 12, or 16 parents.

(4) Comparisons, somewhat limited by the experimental design, gave results conforming to the hypothesis (Robertson 1961) that the inbreeding effect is larger than the amount calculated from population size when both selection intensity and heritability are high.

(5) The effect of population size on the mean was of major importance, relative to the force of selection, only in populations possessing considerable amounts of variation due to dominance effects, their epistatic interactions or both, a result in accordance with the equation proposed by Kempthorne (1957).

(6) Other epistatic variation tended to obscure any differences in means between populations differing in size.

(7) The additive-by-dominance conditional epistatic model produced a situation in which the mean is uplifted rather than depressed by inbreeding.

In general, the results of this study fit the existing theory rather well. It is hoped that clarification of some implications concerning population size and the relation of its effects to specific genetic models and epistasis, in general, has been achieved.

VII. Acknowledgments

The author wishes to express his gratitude to Dr. L. N. Hazel, Department of Animal Science, and Drs. O. Kempthorne and D. L. Harris, Statistical Laboratory, Iowa State University, for their timely counsel during the course of the investigation and preparation of the manuscript; to Mr. K. Rowe for assistance in programming the computer; and to Dr. B. Harshbarger, Head, Department of Statistics, Virginia Polytechnic Institute, for granting the use of time and facilities required in drafting the paper in its final form.

VIII. References

Baker, L. H. (1961).—Linkage and heterozygosity in finite populations. Ph.D. Dissertation, University of Minnesota, Minneapolis.

Barker, J. S. F. (1958*a*).—Simulation of genetic systems by automatic digital computer. III. Selection between alleles at an autosomal locus. *Aust. J. Biol. Sci.* **11**: 603–12.

Barker, J. S. F. (1958*b*).—Simulation to genetic systems by automatic digital computers. IV. Selection between alleles at a sex-linked locus. *Aust. J. Biol. Sci.* **11**: 613–25.

Bennett, J. (1960).—A comparison of selective methods and a test of the pre-adaptation hypothesis. *Heredity* **15**: 65–77.

Bohidar, N. R. (1960).—The role of sex-linked genes in quantitative inheritance. Ph.D. Dissertation, Iowa State University, Ames.

COCKERHAM, C. C. (1954).—An extension of the concept of partitioning hereditary variance for analysis of covariances between relatives when epistasis is present. *Genetics* **39**: 859–82.

CROSBY, J. L. (1960).—The use of electronic computation in the study of random fluctuations in rapidly evolving populations. *Phil. Trans. Roy. Soc.* B **242**: 551–73.

EVANS, G. W., and PERRY, C. L. (1961).—"Programming and Coding for Automatic Digital Computers." (McGraw-Hill Book Company: New York.)

FALCONER, D. S. (1960).—"Introduction to Quantitative Genetics. (The Ronald Press Co.: New York.)

FISHER, R. A. (1918).—The correlation between relatives on the supposition of Mendelian inheritance. Part II. *Trans. Roy. Soc. Edin.* **52**: 399–433.

FRASER, A. S. (1957*a*).—Simulation of genetic systems by automatic digital computers. I. Introduction. *Aust. J. Biol. Sci.* **10**: 484–91.

FRASER, A. S. (1957*b*).—Simulation of genetic systems by automatic digital computers. II. Effects of linkage on rates of advance under selection. *Aust. J. Biol. Sci.* **10**: 492–9.

FRASER, A. S. (1960*a*).—Simulation of genetic systems by automatic digital computers. V. Linkage, dominance, and epistasis. In "Biometrical Genetics". (Ed. O. Kempthorne.) pp. 70–83. (Pergamon Press: New York.)

FRASER, A. S. (1960*b*).—Simulation of genetic systems by automatic digital computers. VI. Epistasis. *Aust. J. Biol. Sci.* **13**: 150–62.

FRASER, A. S. (1960*c*).—Simulation of genetic systems by automatic digital computers. VII. Effects of reproductive rate and intensity of selection on genetic structure. *Aust. J. Biol. Sci.* **13**: 344–50.

GREENBERGER, M. (1961).—Notes on a new pseudo-random number generator. *J. Ass. Comput. Mach.* **8**: 163–7.

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

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

KOJIMA, K. (1961).—Effects of dominance and size of population on response to mass selection. *Genet. Res.* **2**: 177–88.

LEHMER, D. H. (1951).—Mathematical methods in large scale computing units. *Ann. Computn. Lab. Harv.* **26**: 141–5.

LEWONTIN, R. C., and DUNN, L. C. (1960).—The evolutionary dynamics of a polymorphism in the house mouse. *Genetics* **45**: 705–22.

LUSH, J. L. (1948).—"The Genetics of Populations." Mimeographed book. (Department of Animal Science, Iowa State University, Ames.)

MALÉCOT, G. (1948).—"Les Mathématiques de l'Hérédité." (Masson et Cie.: Paris.)

MARTIN, F. G. JR., and COCKERHAM, C. C. (1960).—High speed selection studies. In "Biometrical Genetics". (Ed. O. Kempthorne.) pp. 35–45. (Pergamon Press: New York.)

MULLER, M. E. (1959).—A comparison of methods for generating normal deviates on a computer. *J. Ass. Comput. Mach.* **6**: 376–83.

PEACH, P. (1961).—Bias in pseudo-random numbers. *J. Am. Statist. Ass.* **56**: 610–18.

ROBERTSON, A. (1960).—A theory of limits in artificial selection. *Proc. R. Soc.* B**153**: 234–49.

ROBERTSON, A. (1961).—Inbreeding in artificial selection programmes. *Genet. Res.* **2**: 189–94.

ROTENBERG, A. (1960).—A new pseudo-random number generator. *J. Ass. Comput. Mach.* **7**: 75–7.

TAUSSKEY, O., and TODD, J. (1956).—Generation of pseudo-random numbers. In "Symposium on Monte Carlo Methods". (Ed. H. A. Meyer.) (John Wiley and Sons, Inc.: New York.)

WRIGHT, S. (1931).—Evolution in mendelian populations. *Genetics* **16**: 97–159.

WRIGHT, S. (1935).—The analysis of variance and the correlation between relatives with respect to deviations from an optimum. *J. Genet.* **30**: 243–56.

WRIGHT, S. (1951).—The genetic structure of populations. *Ann. Eugen.* **15**: 323–54.