A MONTE CARLO EVALUATION OF PREDICTED SELECTION RESPONSE*

By J. L. GILL†

[*Manuscript received March* 16, 1965]

Summary

This study utilizes a computer programme that simulates the effects of selection, linkage, and environmental variation on the genetic progress of finite populations. The results, with respect to selection advance and the inbreeding that occurs in small populations, have been discussed previously (Gill 1965). In the present paper, the correspondence of Monte Carlo results to the hypothetical progress predicted according to a mathematical formulation given by Griffing (1960) is evaluated.

The predicted contribution to change in the mean attributed to additive-by-additive genetic variance, in most cases, was far too large over several generations of selection, or for even shorter periods in the smallest populations. Random genetic drift and selection appeared to have considerable influence in changing the genetic parameters quickly. The effects of restricted population size and selection on changes in value of genetic parameters and the effects peculiar to a particular mode of gene action combined to obscure the prediction problem so that Griffing's theoretical expression was accurate for more than a very few generations only when a fortuitous combination of several factors occurred. However, the magnitude of the discrepancies noted between predicted means and those observed in Monte Carlo populations possibly is larger than it would be in a practical situation because of restrictions in the mechanics of simulation.

I. THEORY OF PREDICTING RESPONSE TO SELECTION

Kempthorne (1957, ch. 16) summarized much of the theory concerning changes in the mean of non-random-mating diploid populations under selection, and showed that the change in the mean could be expressed in terms of changing values of the ratio of heterozygote frequency to the product of homozygote frequencies.

Kimura (1958) described the rate of increase in population fitness for an arbitrary number of loci, and adapted the formula to artificial selection by utilizing heritability and selection differential to account for additive effects, relating dominance deviations to changes in gene frequency and the coefficient of inbreeding and, to account for epistatic deviations, by weighting with coefficients of departure from random mating, which is similar to the method used by Fisher (1941) and extended by Kempthorne (1957). Kimura's formula possibly is the most comprehensive yet developed for descriptive purposes.

* Journal Paper No. J-5073, 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, Iowa; present address: Department of Dairy, Michigan State University, East Lansing, Michigan.

Kojima (1961) developed an expression for the effects of dominance and size of population on response to mass selection for the single-locus case. Generalization to the multiple-locus case is possible only when epistasis and linkage disequilibrium are not present, and terms for dominance effects appear to be less general than Kimura's.

Griffing (1960) developed an argument for the general problem of prediction associated with truncated selection based on the individual phenotype, with arbitrary dominance, epistasis, linkage, and number of loci. His analysis proceeds on the level of the gene, the gamete, and the genotype, and requires four major assumptions:

(1) Populations of infinitely large size are required to completely exclude error due to the drift of gene frequency.

(2) The quantity (gene effect ÷ phenotypic standard deviation) must be small enough so that the squares of such quantities are negligible in magnitude, and may be ignored.

(3) All selected genotypes should have the same reproductive value (to exclude differential natural selection).

(4) Recombination frequency should be the same in both sexes.

The analysis based on genes is analagous to the standard prediction equation, additive heritability times selection differential, and implies that the population mean will not degenerate if selection is suspended. The gametic analysis of genotypes generated by alleles at two loci that may be linked follows the generalized gene model of Kempthorne (1957, p. 416) and the partition of genotypic variance into additive, dominance, and individual epistatic components.

Griffing derived the predicted change in the mean after n generations of selection as

$$\Delta\mu_n = (I/\sigma_p^2)[n\sigma_A^2 + \sum_{i=1}^{n}(1-r)^{i-1}\tfrac{1}{2}\sigma_{AA}^2],$$

where I is the selection differential, r is the recombination frequency between the two loci, and σ_p^2, σ_A^2, and σ_{AA}^2 are total phenotypic, additive, and additive-by-additive epistatic variance components, respectively. If the loci exhibit a low r, considerable effect can be generated by σ_{AA}^2, but if the loci are independent, the maximum contribution after n generations of selection is approximately σ_{AA}^2, for large n. The influence of σ_{AA}^2 diminishes as the number of generations in the selection programme increases. That causes the departure from linearity of the response to selection with time. Thus, the long-time response to selection should result in an asymptotic approach to the goal of selection, whether it is a goal of homozygosity or of stable equilibrium. Griffing also showed that the degeneration of the mean occurring upon relaxation of selection for m generations is due to slippage of the contribution of σ_{AA}^2 to a fraction $(1-r)^m$ of its original value.

Griffing's third analysis, based on individual genotypes, relates mean progress to the covariance of parent and offspring, and leads to predictions analogous to those of the analysis based on gametes.

II. Experimental Method

A comparison of the results observed for simulated populations involving dual epistasis and the results predicted by Griffing's (1960) method was designed to show the nature of some of the limitations of the theoretical expression. The assumption that the genetic parameters are constant under selection is as likely to lead to serious discrepancies between observed and predicted results as is the failure to account for the random drift that occurs in finite populations. Furthermore, the errors introduced by approximations tend to accumulate, so that the basis of prediction becomes more subject to error as the mean of the selected population becomes further removed from its original position. However, simulation results will not provide complete assessment of the bias in Griffing's (1960) equation as applied to small populations. Limitation of simulation to a relatively few loci may cause the magnitude of gene effects, relative to total variability, to be unrealistically large, and changes in population parameters with selection may be distorted relative to time.

Populations of unisexual diploid individuals were simulated, and their quantitative characteristics were assumed to be expressed in both sexes. Population parameters that were varied include number of parents, selection intensity, linkage, and environmental variation. A metric characteristic was determined by genes at 40 loci equally spaced over eight chromosomes, with two alleles per locus and equal genetic effects at all loci. Equal numbers of parents of each sex, selected by upper truncation of ranked phenotypes over a period of 30 generations, were mated at random by sampling with replacement, which allows for full-sibs and half-sibs among the progeny. Linkage of adjacent loci (without interference) was simulated. The initial parent population was completely heterozygous, with random association of coupling and repulsion linkage phases. Monte Carlo procedures for random number generation were utilized in simulating the probabilistic genetic mechanisms.

Parameters for particular populations were chosen from combinations of parent populations of 32, 16, 12, or 8 individuals, selection intensties of $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{6}$, or $\frac{1}{8}$, recombination fractions of 0·5, 0·2, 0·05, or 0·005, and initial genotypic fractions of phenotypic variance of 1, $\frac{3}{4}$, $\frac{1}{2}$, or $\frac{1}{4}$. The experimental design involved replication of a fractional sixteenth of this 4^4 factorial plan for each genetic model. A more detailed description of the mechanics of simulation and the experimental design has been given by Gill (1965).

This study concerns the results from four genetic models: additive, complete dominance, optimum gene number (Wright 1935), and additive-by-additive ($A \times A$) conditional epistasis (Gill 1965). The latter two models involve epistasis and may be represented by genotypic values for two interacting loci as follows:

	Optimum Number				$A \times A$		
	AA	*Aa*	*aa*		*AA*	*Aa*	*aa*
BB:	7	10	11	*BB*:	12	10	8
Bb:	10	11	10	*Bb*:	10	10	10
bb:	11	10	7	*bb*:	8	10	12

The predictions for additive and dominance cases are the usual linear ones, because neither of those models involves the additive-by-additive variance (σ^2_{AA}) included in

the second term of Griffing's (1960) prediction equation, but the results give indications of the type and severity of errors caused by some to the assumptions, and are of interest as a check on the standard linear prediction. Predictions were based on two

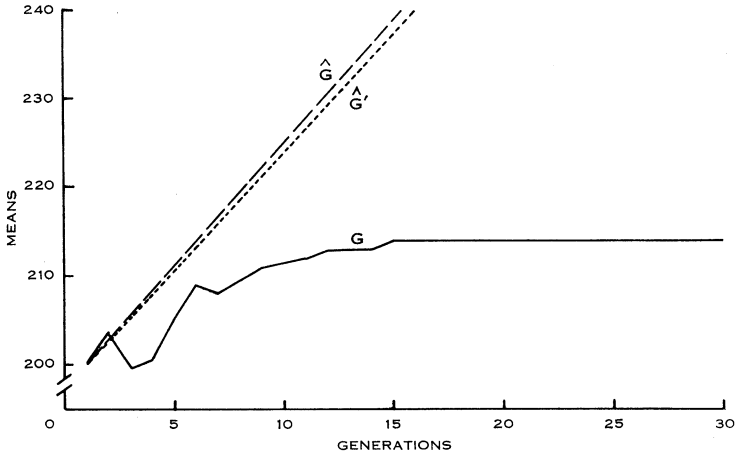

Fig. 1.—Relation of the mean (G) of a small population ($n = 8$) to that predicted from parameters of simulation ($\hat{G}$) and from parameters observed in the first generation of selection ($\hat{G}'$) with additive gene action.

methods for each population: (1) the parameters that were used to simulate the original population, and (2) parameters measured in an early generation of the Monte

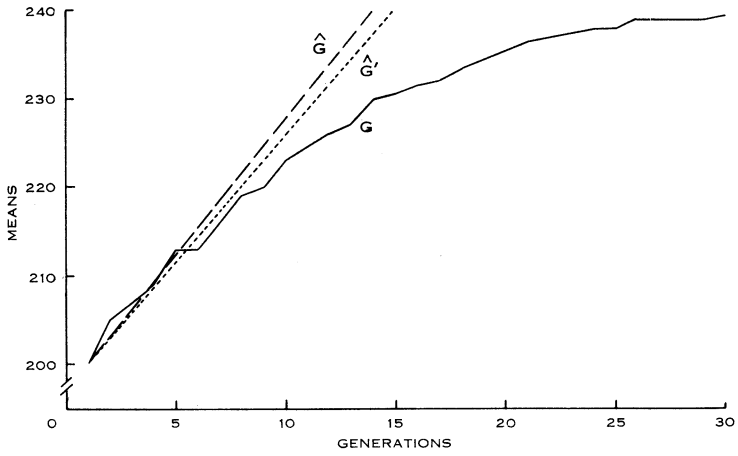

Fig. 2.—Relation of the mean (G) of a "large" population ($n = 32$) to that predicted from parameters of simulation ($\hat{G}$) and from parameters observed in the first generation of selection ($\hat{G}'$) with additive gene action.

Carlo selection results. The first method involves some conditions that are peculiar to the original population (e.g. gene frequency of 0·5 at each locus). The second method circumvents that problem and provides a measure of bias caused by the

change in parameters from their original values. The accompanying figures give comparisons of three results:

(1) $\hat{G}$, the mean predicted by Griffing's equation using the first method.

(2) $\hat{G}'$, the mean predicted using the second method.

(3) G, the observed mean of simulated populations.

III. Monte Carlo Results

For the additive model, Figures 1 and 2 illustrate results for two parent populations restricted to 8 and 32 individuals, respectively. Two major conclusions stand out:

(1) The differences between $\hat{G}$ and $\hat{G}'$ are not significant ($\hat{G}'$ based on generation 1 parameters). The implication is that genetic parameters changed very little in one generation.

(2) The large differences between predicted and observed results after two generations in the smaller population, and after eight generations in the larger one, emphasize the seriousness of Griffing's assumption of infinitely large populations.

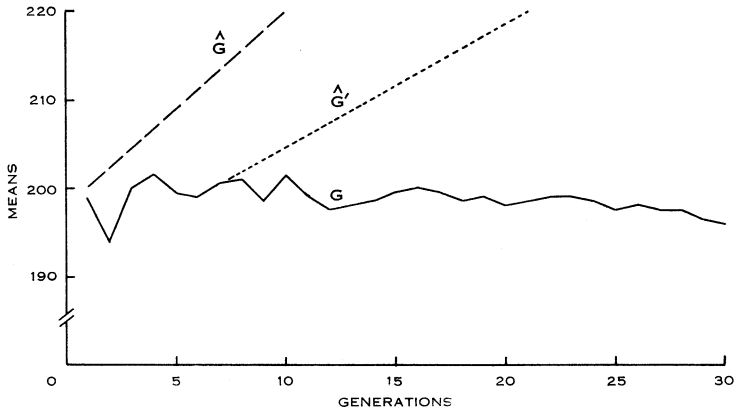

Fig. 3.—Relation of the mean (G) of a small population ($n = 8$) to that predicted from parameters of simulation ($\hat{G}$) and from parameters observed in the sixth generation of selection ($\hat{G}'$) with complete dominance.

For the complete dominance model, $\hat{G}'$ was calculated from parameters observed in generation 6, so that the effect of changing parameter values might be assessed. One would presume that $\hat{G}'$ should be a superior predictor to $\hat{G}$, but that proved to be true consistently only for the smallest populations, and neither predictor was particularly good for more than two or three generations because of inbreeding depression. Figure 3 illustrates an extreme case where the parent population was restricted to eight individuals. The genetic parameters did change somewhat in six generations in the smaller populations (note slope difference for $\hat{G}$ and $\hat{G}'$) but changed very little in that period in the larger populations.

To test Griffing's theory about the effects of linkage and epistasis on prediction, one must utilize models that involve additive-by-additive variation (σ^2_{AA}) in more

than trivial amounts. Wright's optimum gene number model and the additive-by-additive ($A \times A$) conditional epistastic model (previously described) meet this criterion.

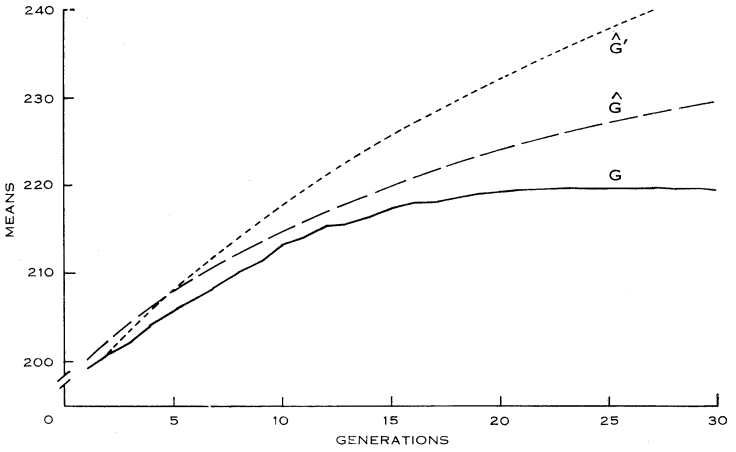

Fig. 4.—Relation of the mean (G) of a "large" population ($n = 32$) to that predicted from parameters of simulation (G) and from parameters observed in the second generation of selection ($\hat{G}'$) with limited recombination (0·05) under the optimum gene number model.

For the optimum model, σ^2_{AA} represents two-thirds of the total genotypic variance when gene frequencies are at 0·5, and nearly one-third for frequencies of 0·9 or 0·1

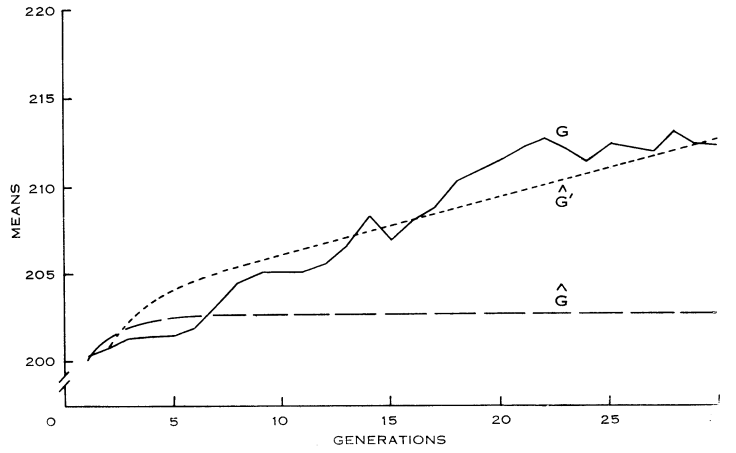

Fig. 5.—Relation of the mean (G) of a "large" population ($n = 32$) to that predicted from parameters of simulation ($\hat{G}$) and from parameters observed in the second generation of selection ($\hat{G}'$) with free recombination under the optimum gene number model.

(Gill 1965). The $A \times A$ model is defined as involving only σ^2_{AA} genetic variance when gene frequencies are at 0·5, but the ratio of σ^2_{AA} to total genotypic variance decreases as gene frequency changes, σ^2_A becoming more important.

Some representative results for the optimum model are shown by Figures 4 and 5. Three results stand out:

(1) Random drift in the smaller populations contributed heavily to the discrepancy between predicted and observed results.

(2) Both predictors overestimated progress, sometimes considerably, when the amount of recombination was limited to 0·05 or less.

(3) When linkage was reasonably close, prediction based on simulated parameters ($\hat{G}$) was better than that based on parameters observed in the second generation ($\hat{G}'$), but the reverse was more usually true for low levels of linkage or free recombination.

Most of the overprediction when recombination was quite limited was due to the decline in σ^2_{AA} from the amount used in the prediction equations. With free recombination, the maximum theoretical contribution of epistasis is small (exactly σ^2_{AA}), that maximum being achieved and remaining constant after 15 generations of selection. Thus, the decline in σ^2_{AA} from the original amount did not contribute as much to accumulated errors in the prediction when there was no linkage.

The discrepancy between the two predictors largely is due to the curious phenomenon that the optimum model does not involve any additive variance (σ^2_A) at gene frequencies of 0·5, but considerable amounts of it are generated by small changes in gene frequency. Therefore, the predictor $\hat{G}$, which utilizes the parameters of simulation of the original heterozygous population, involves no σ^2_A, whereas the predictor $\hat{G}'$ utilizes the amount of σ^2_A generated after two generations of mating and one generation of selection have occurred. That amount varied from 10 to 60% of the total genotypic variance, somewhat inversely with population size, implicating random drift as the main generator of σ^2_A at that time. Because of this situation, $\hat{G}$ was a better predictor than $\hat{G}'$ by default when linkage was tight, even though both grossly overestimated the observed genetic progress because of large predicted contributions of σ^2_{AA} (Fig. 4). However, for low levels of linkage or for free recombination, $\hat{G}$ involved no predicted contribution from σ^2_A and very little from σ^2_{AA}. Therefore, progress was underestimated by $\hat{G}$ after a few generations of selection, while $\hat{G}'$ still overestimated progress, except for some of the larger populations, because of random drift and because the contribution of σ^2_A was extrapolated linearly, not allowing for changing parameters (Fig. 5).

As in the case of the optimum model, prediction under the $A \times A$ conditional epistatic model from simulated parameters, $\hat{G}$, was based solely on the contributions of σ^2_{AA}. Again, additive variance is generated very quickly and in large amounts when gene frequencies change from 0·5 (cf. Gill 1965). Therefore, $\hat{G}$ was somewhat better as a long-term predictor than $\hat{G}'$ (based on parameters observed in generation 6) when linkage was tight, especially for the larger populations (Fig. 6), although there was a tendency in many populations to underestimate progress in the early generations and to overestimate later. However, the fact that $\hat{G}$ was a reasonably good estimator for any cases with tight linkage appears to be purely coincidental, the result of overestimation from assuming constant σ^2_{AA} and underestimation through failing to consider the generation of σ^2_A. This is corroborated by the fact that, when low levels

of linkage or free recombination existed, diminishing the theoretical predicted effect of σ^2_{AA}, $\hat{G}$ badly underestimated the means of all populations after two or three generations of selection (Fig. 7).

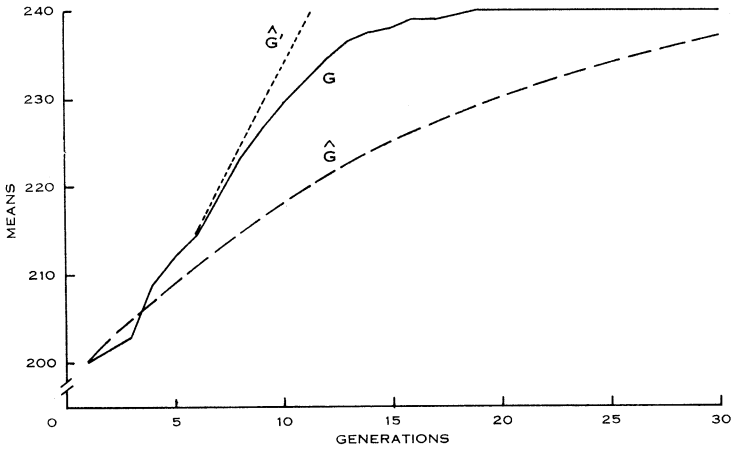

Fig. 6.—Relation of the mean (G) of a "large" population ($n = 32$) to that predicted from parameters of simulation ($\hat{G}$) and from parameters observed in the sixth generation of selection ($\hat{G}'$) with limited recombination (0·05) and additive-by-additive conditional epistasis.

One might expect to make superior predictions by using parameters observed in generation 6 of the Monte Carlo results ($\hat{G}'$) because the development of σ^2_A can be

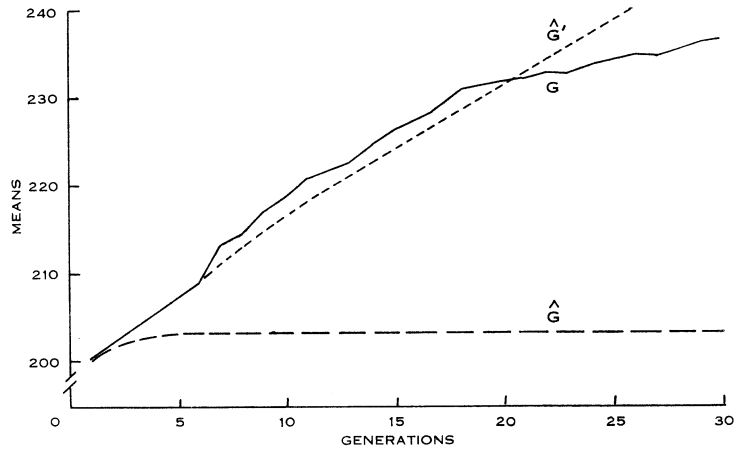

Fig. 7.—Relation of the mean (G) of a "large" population ($n = 32$) to that predicted from parameters of simulation ($\hat{G}$) and from parameters observed in the sixth generation of selection ($\hat{G}'$) with free recombination and additive-by-additive conditional epistasis.

taken into account. However, the predictions were generally above the observed mean of populations with tight linkage because of the decrease of σ^2_{AA} over time. With low levels of linkage or free recombination, prediction with $\hat{G}'$ was relatively

good for several generations because the theoretical effect of σ^2_{AA} was small and was not much affected by changing parameters and because σ^2_A apparently was being generated and utilized at nearly equal rates until several generations of selection forced gene frequencies to more extreme values.

The results emphasize the difficulty of utilizing projections based on infinite population size in realistic populations of restricted size. The futility of predicting for more than a few generations without a re-evaluation of genetic parameters is evident, whether the predictions are linear or asymptotic to the selection goal. Random genetic drift, as well as selection, has considerable influence in changing parameter values rather quickly, in terms of generations. It must be stressed, however, that the dynamics of the simulated populations may differ from the rate and magnitude of change observed in natural populations.

IV. Acknowledgments

The author wishes to express his gratitude to Drs. L. N. Hazel, O. Kempthorne, and D. L. Harris of Iowa State University for their counsel during the investigation and preparation of the manuscript, and to Dr. Boyd Harshbarger, Head, Department of Statistics, Virginia Polytechnic Institute, for granting the use of time and facilities required in drafting the paper.

V. References

Fisher, R. A. (1941).—Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**: 53–63.

Gill, J. L. (1965).—Effects of finite size on selection advance in simulated genetic populations. *Aust. J. Biol. Sci.* **18**: 599–617.

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

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.

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

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