# SIMULATION OF GENETIC SYSTEMS BY AUTOMATIC DIGITAL COMPUTERS

## VI. EPISTASIS

By A. S. Fraser*

*Summary*

Simulation, by Monte Carlo methods, of the effect on the genotype of selection against phenotypic extremes has shown that selection will lead to fixation of a simple additive genetic system at an extremely slow rate in all but very small populations. In complex epistatic systems, such selection operates to modify the relation of the genotype to the phenotype. The relationship becomes an S-shaped function. The efficiency of selection is independent of population size. The deviation from initial gene frequencies due to selection is far less per unit decrease of phenotypic variability in the epistatic than in the additive lines.

## I. Introduction

The work of this Laboratory on quantitative inheritance has recently centred on the problems posed by the existence of genetic variability in the absence of phenotypic variability (see Rendel 1959; Fraser and Kindred 1960). Consequently, the work with simulation programmes on electronic computers has been directed at the problem of selection against extreme phenotypes. Robertson (1956), in a theoretical analysis of this problem, examined two alternatives: the first, in which extremes are less fit because they are extremes, he found would lead to fixation; the second, in which extremes are less fit not because they are extremes but because they are homozygotes, he found would maintain genetic variability. He, thus, clearly supports Lerner's thesis that heterozygotes have an enhanced fitness *per se*, i.e. that there is overdominance of fitness (see Lerner 1958). Robertson did not consider the effect of such selection against extreme deviants on epistatic interactions, yet it seems probable that selection can modify such interactions and, consequently, reduce the trend towards homozygosity (Rendel 1959). A programme written to simulate a genetic system in which both the dominance and epistatic relations were under genetic control indicated that fixation of the basic loci will be considerably reduced by modification of epistatic relations (Fraser 1959). There were, however, several features of this programme which might produce bias, the most important being an inherent asymmetry. In addition, the programme was, by computer standards, slow and it was limited to a maximum population size of slightly less than 200 progeny. A new programme has been written which, although based on a larger number of loci—40 as against 20—is faster and puts no limit on the number of progeny per generation, the limit being on the number of parents. This programme is described below, and results from two sets of its runs discussed.

* Division of Animal Genetics, C.S.I.R.O., University of Sydney.

## II. Structure of the Epistasis Programme

The programme has six main sections:

(1) Extract, without replacement, a pair of parents at random from the given set of parents.

(2) Form a set of progeny from these parents.

(3) Determine the phenotypes of the progeny.

(4) Select potential parents from the set of progeny.

(5) Repeat (1)–(4) until all parents have produced the specified number of progeny.

(6) Print out any required information.

(7) Repeat (1)–(6) using the selected progeny as parents.

## III. Simulation of Segregation

Section II includes the simulation of segregation.   The method initially described by Fraser (1957a, 1957b) is based on a "random walk" along the length of the genotype, each locus being considered separately. Mr. J. B. Butcher, Adolph Basser Computing Laboratory, has suggested a method which obviates considering each locus in sequence, and makes full use of the parallel arithmetic of the SILLIAC computer.

A single SILLIAC register contains 40 "bits" of binary information, and orders in the SILLIAC suppressing carry-over allow various operations of "logical algebra" to be performed separately, and simultaneously, for all 40 bits of a register. From such operations it is possible to identify the genetic status of 40 loci in a single process.   Given such identification it is then easy to simulate segregation simultaneously at 40 loci.

This is accomplished by the following sequence, given that register A contains the "maternal genotype" and register B the "paternal genotype" of a parent which is to produce gametes:

(1) Form the logical product (L.P.) of A and B

$$
\begin{array}{ccccccc}
1 & 1 & 0 & . & . & . & . & . & A \\
1 & 0 & 0 & . & . & . & . & . & B \\
\hline
1 & 0 & 0 & . & . & . & . & \text{L.P. (A/B)}
\end{array}
$$

(2) Form the logical non-equivalent (L.N.E.) of A and B

$$
\begin{array}{ccccccc}
1 & 1 & 0 & . & . & . & . & . & A \\
1 & 0 & 0 & . & . & . & . & . & B \\
\hline
0 & 1 & 0 & . & . & . & . & \text{L.N.E. (A/B)}
\end{array}
$$

The L.P. (A/B) identifies the loci homozygous for the 1 type alleles; the L.N.E. (A/B) identifies the heterozygous loci.

(3) Form the L.P. of a random number (C) with the L.N.E. (A/B)

|   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | . | . | . | . | . | L.N.E. (A/B) |
| 0/1 | 0/1 | 0/1 | . | . | . | . | . | C |

| 0 | 0/1 | 0 | . | . | . | . | . | L.P. (L.N.E. (A/B)/C) |

(4) Add the L.P. (A/B) to the L.P. (L.N.E. (A/B)/C)

|   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| 0 | 0/1 | 0 | . | . | . | . | . | L.P. (L.N.E. (A/B)/C) |
| 1 | 0 | 0 | . | . | . | . | . | L.P. (A/B) |

| 1 | 0/1 | 0 | . | . | . | . | Simulated gamete |

As a result of these operations 1 is placed in every position representing a gene of the 1 type which was homozygous in the parent and 0 in every position representing a homozygous gene of the 0 type; the heterozygous genes are represented by 1 or 0 at random.

These operations can be performed simultaneously for up to 40 digits. Apart from the formation of a random number, this sequence takes seven orders, totalling 395 $\mu$sec.

### IV. DETERMINATION OF THE PHENOTYPE

The genetic model is of 40 independent loci. The phenotype determined by these loci has two components, $P_0$ and $P_1$, each determined similarly by 20 loci. Each group of 20 loci is considered as four sets of five loci. The relationships of these to the phenotype are the same as those in the initial epistasis programme (Fraser 1959). These four subgenotypes are termed the $A$, $D$, $Q$, and $C$ genotypes respectively.

The $A$ group of five loci determines a contribution to the phenotype which in the absence of dominance or epistasis is given for each locus by

| Genetic Status | Phenotypic Contribution |
|:---:|:---:|
| $\frac{1}{1}$ | $-1\cdot0$ |
| $\frac{1}{0}$ | $0$ |
| $\frac{0}{0}$ | $+1\cdot0$ |

The "additive" contribution of the $A$ subgenotype to the phenotype is given by summation over the five loci.

The $D$ group of five loci determines the degree of dominance at the loci of the $A$ genotype. This genotype simulates a dominance modifier system affecting identically the five loci of the $A$ genotype. It determines the phenotypic contribution of heterozygotes of the $A$ group, and has no effect on the phenotypic contribution of the homozygotes.

The $D$ loci do not themselves have any dominance, i.e. over the five loci of this genotype there are 11 possible values conferring dominance on the $A$ group. A vector of possible dominance values is specified for a particular run, and remains constant for that run. The value of the $D$ genotype in an individual determines which value of this vector is appropriate.

If the vector of possible dominance coefficients is $\{d_i\}$, then:

| Genetic Status | Phenotypic Contribution |
|---|---|
| $\dfrac{1}{1}$ | $-1 \cdot 0$ |
| $\dfrac{1}{0}$ | $d_i$ |
| $\dfrac{0}{0}$ | $+1 \cdot 0$ |

The vector $\{d_i\}$ has been set, for the runs which are discussed in this paper, at either zero, or in a linear sequence covering the range $+1 \cdot 0 \leqslant d_i \leqslant -1 \cdot 0$, such that a substitution in the $D$ genotype of a 0 for a 1 type allele makes a difference of $0 \cdot 2$ in $d_i$, which gives a variation from complete dominance of the 1 alleles ($d_i = -1 \cdot 0$) through no dominance ($d_i = 0 \cdot 0$) to complete dominance of the 0 alleles ($d_i = +1 \cdot 0$).

Summation over the five loci of the $A$ genotype then gives the additive + dominance contribution to the phenotype. This is termed $(P_a + P_d)_0$.

The degree of interaction between the five loci of the $A$ genotype is determined by the $Q$ and $C$ genotypes. Just as a particular $D$ genotype determines a specific value of $d_i$, so the $Q$ and $C$ genotypes determine specific values of quadratic and cubic interaction coefficients from the vectors of such coefficients: $\{q_i\}$ and $\{c_i\}$. These have been set for the runs which are discussed in this paper at either zero, or in linear sequence covering the following ranges:

$$+0 \cdot 5 \leqslant q_i \leqslant -0 \cdot 5; \quad +0 \cdot 125 \leqslant c_i \leqslant -0 \cdot 125.$$

The coefficients $q_i$ and $c_i$, determined by specific $Q$ and $C$ genotypes, give the degree of interaction by

$$P_Q = (P_a + P_d)_0^2 \, q_i,$$

and

$$P_C = (P_a + P_d)_0^3 \, c_i.$$

The complete determination of the $P_0$ component of the phenotype is given by

$$P_0 + (P_a + P_d)_0 + q_i(P_a + P_d)_0^2 + c_i(P_a + P_d)_0^3.$$

The other 20 loci similarly determine $P_1$, and the total phenotype is the simple sum of $P_0$ and $P_1$.

The subdivision of the phenotype into two independently determined components was introduced to allow independent variation of the dominance genotypes. Mather (1943) has suggested that the lack of dominance in a quantitative system

may be due to individual loci of the system having dominance values in opposite directions. The average value of dominance over all the loci can then be zero, and variation still have a dominance component. In our model, two values of dominance are possible, one for each set of $A$ loci, and it was originally considered that this could provide data on the feasibility of Mather's suggestion. Latter (personal communication) has made the point that each $D$ genotype sets a dominance value for five $A$ loci, and, consequently any trends towards different dominance values between these loci would average out to zero. An adequate test of Mather's hypothesis requires an individual dominance modifier system for each $A$ locus. The separation of the phenotype into two components is therefore only of value in indicating the consistency of any trends.

## V. Method of Selection

In this programme, each progeny as it is formed is tested against fixed phenotypic limits which are specified at the beginning of the run, and remain constant for the duration of the run. Since the number of individuals required as parents $(n)$ may be less than the number of individuals with phenotypes within limits, the first $n$ acceptable parents are provisionally accepted and placed in the "parent" store. Thereafter each acceptable parent is tested for substitution in that store against a random number, with a probability of $n/i$, where $i$ is the total number of acceptable parents which have so far been tested. This ensures that all potential parents have an equal probability of being accepted as parents. The selection limits set for the runs discussed in this paper are $\pm 1 \cdot 0$, i.e. the effect of a single gene substitution on the simple additive scale.

This method of simulating selection, although fast, has the disadvantage that the selection limits cannot be easily varied during a run. Another method, which will be used in a modification of this programme, does not have this disadvantage. It is based on the sequence of pseudo-random numbers being completely determined. Given the values of the random numbers at a specific point in a run, it is possible, by substituting these values, to repeat any particular sequence of the operations of the programme. This ability to repeat a sequence can be used to determine selection limits. At the beginning of the formation of progeny the values of the random number generator are stored. Then all, or a sample, of the progeny are formed and their phenotypes computed. Only sufficient information is retained to describe the frequency distribution of phenotypes. From this, selection limits are computed which will contain a specified fraction of the progeny. Given such limits, the original values of the random number generator are substituted, and the generation of progeny restarted. Each progeny as formed is then tested for acceptance against the selection limits.

## VI. Pre-set Parameters

Although the "epistasis" programme has been constructed to minimize the number of parameters initially specified, and thereafter, maintained constant, some must be specified and these are:

(1) Whether random mating or self-fertilization is to occur.

(2) The number of parents.

(3) The number of progeny per mating.

(4) The maximum and minimum limits of eligible phenotypes.

(5) The identification number of the run.

(6) The vectors of dominance and interaction coefficients.

(7) The genotypes of the initial set of parents.

Sixteen runs of the programme have been made: eight in the absence of any dominance or epistasis, and eight in the presence of such effects. These runs were made at four population sizes: 20, 40, 80, and 160 parents, two runs at each population size. The number of progeny per mating was set at 50 for all runs, making the total numbers of progeny 500, 1000, 2000, and 4000. Two independent runs were made for each set of parameters.
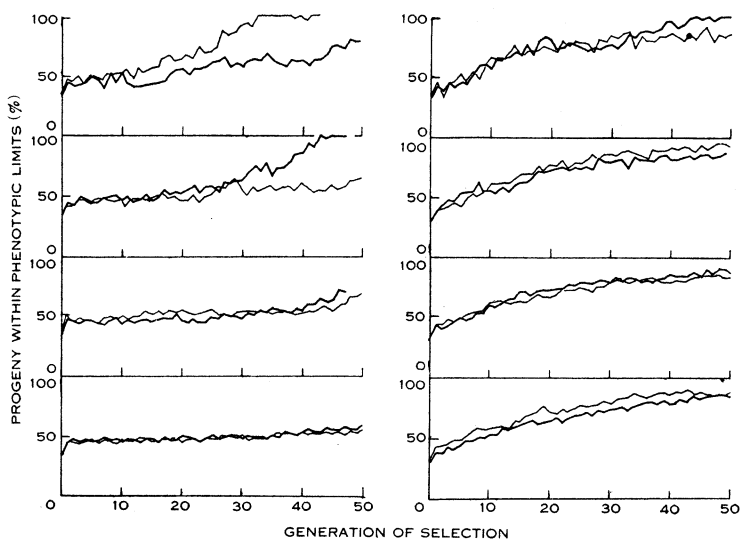


Fig. 1.—Percentage of progeny within the phenotypic limits plotted against generation of selection. Runs made in the absence of dominance and epistasis on the left. Runs made in the presence of dominance and epistasis on the right.

The absence of dominance and epistasis was produced in the first eight runs by setting $\{d_i\}$, $\{q_i\}$, and $\{c_i\}$ at zero. These genotypes consequently have no effect on the phenotype, and selection is directed solely at the $A$ genotype, i.e. selection is operating on a simple additive system of 10 loci. The $D$, $Q$, and $C$ loci, totalling 30, are, in these runs, not under selection and changes of their frequencies can be taken as the basis for comparison with the $A$ loci to determine the changes produced by selection.

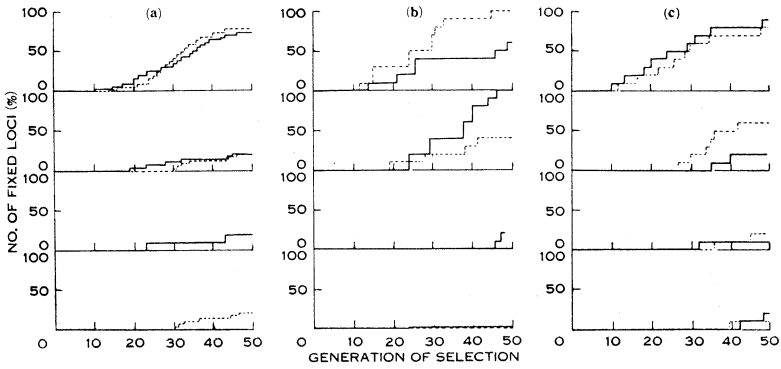All parents at the beginning of all runs were heterozygous at all loci.

Fig. 2.—(a) Genetic fixation in the absence of selection. The percentage of *D*, *Q*, and *C* loci which have become fixed is plotted against generation for the eight runs in which these loci have no phenotypic effect. (b) Genetic fixation of a simple additive system. The percentage of *A* loci which have become fixed, plotted against generation of selection for the first eight runs. (c) Genetic fixation of the *A* loci in runs which include selection for modifiers of dominance and epistasis.
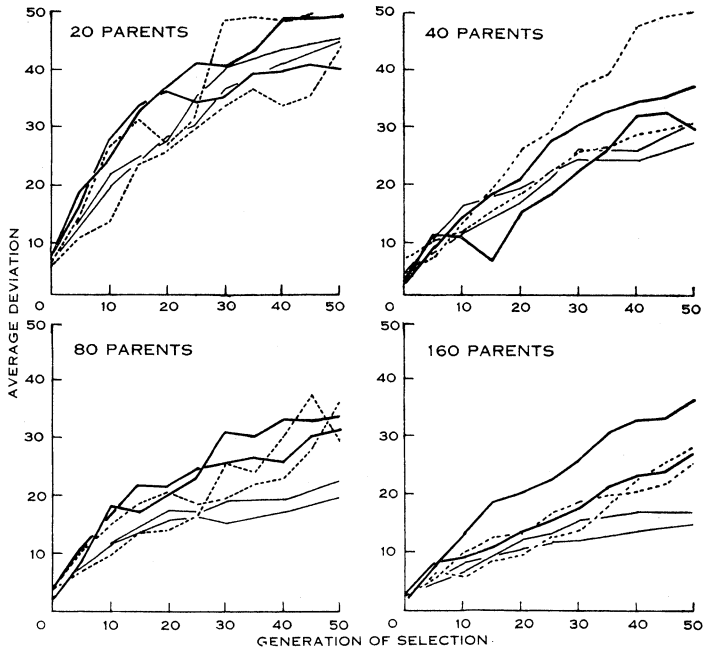


Fig. 3.—Average deviation of the frequencies of the *A* loci from the initial value of 0·5 is shown plotted against generation for (a) unselected loci (- - - -); (b) simple additive system (——); (c) epistatic system (——).

## VII. RESULTS AND DISCUSSION

### (a) Phenotypic Variability

The percentage of progeny produced each generation which have phenotypes within the specified phenotypic limits is a measure of the effectiveness of selection against the extreme phenotypes. These percentages are shown plotted against generation of selection in Figure 1.

In considering the eight runs made in the absence of dominance and epistasis, there is a marked effect of population size. In small populations, selection causes,
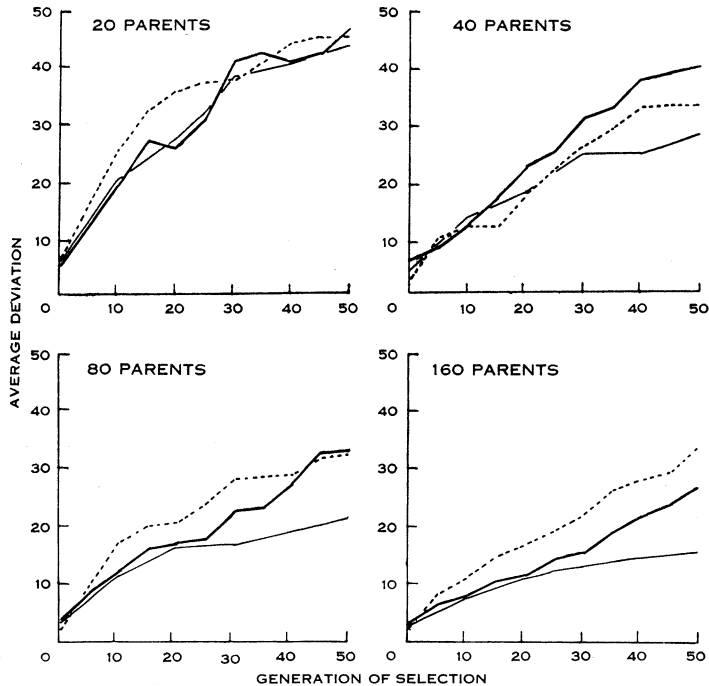


Fig. 4.—As for Figure 3, but the two replicate runs have been averaged.

at the small population sizes, a marked increase of the percentage of progeny with phenotypes within the selection limits. In large populations the percentage of acceptable progeny increases more slowly. This contrasts very markedly with the eight runs made in the presence of dominance and epistasis; the percentage of "acceptable" progeny increases rapidly over the initial generations of selection, and then more slowly as selection proceeds. The effect of population size is slight, being most apparent in a lower variation from generation to generation in larger populations (see Fig. 1).

The effect of selection on phenotypic variability in the two sets of populations is sufficiently different to suggest that different mechanisms are operating. This is shown by the genotypes which the machine recorded each generation.

The effect of selection and population size on the distribution of gene frequencies is shown in Figure 2 in which the percentage of loci which have become fixed is plotted against generation of selection and in Figure 3 in which the average deviation of gene frequencies from the initial value of $0 \cdot 5$ is shown plotted against generation of selection.

The incidence of genetic fixation shows that there is a marked effect of population size and that there is little if any effect of selection. The deviations from the original gene frequency show similarly that there is a marked effect of population size. Selection is ineffective in smaller populations of 20 and 40 parents but in ones
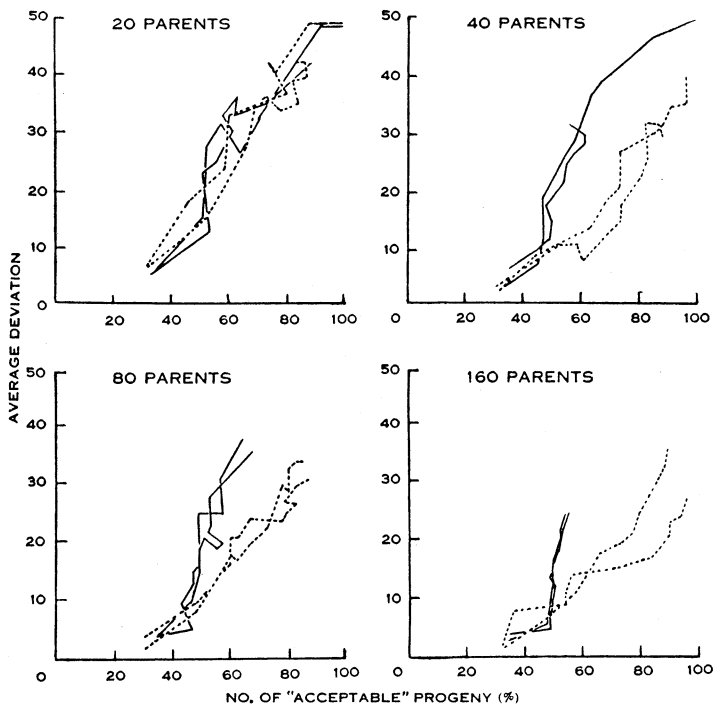


Fig. 5.—Average deviation of the frequencies of the *A* loci from initial value of $0 \cdot 5$ plotted against the percentage of "acceptable" progeny.
———— Additive system; – – – – epistatic system.

of 80 and markedly in ones of 160 parents there is a real difference between the changes of gene frequency of unselected and selected loci. This is shown more clearly in Figure 4, in which the two replicate runs have been averaged.

Figure 4 indicates that selection of an epistatic system produces a greater change of gene frequency than selection of a simple additive system. However, no account has been taken of the effectiveness of selection in reducing phenotypic variability. This is particularly evident in large populations. In large populations without dominance or epistasis, i.e. a simple additive system, selection for over 50 generations has produced only a slight reduction of phenotypic variability whereas with dominance or epistasis, i.e. for the epistatic system, selection has

produced a very marked reduction of variability. Consequently, the percentage of progeny within limits is shown plotted against deviations from the initial gene frequency in Figure 5. These graphs show that selection against extreme phenotypes produces a much greater reduction of variability for a specified deviation from the initial gene frequency in the "epistatic" runs than in the "additive" runs.

The dominance and epistasis runs were made with $\{d_i\}$, $\{q_i\}$, and $\{c_i\}$ set to the ranges specified in Section IV. Variation of the $D$, $Q$, and $C$ genotypes due to segregation will therefore produce variation of $d_i$, $q_i$, and $c_i$. This will affect $P_i$ and consequently selection can affect the $D$, $Q$, and $C$ genotypes. The mean values



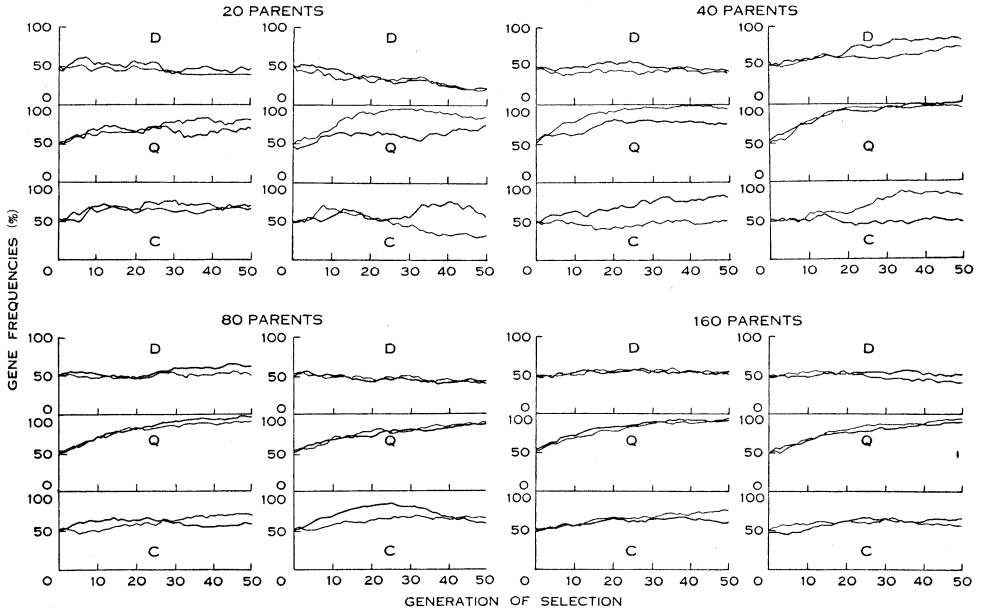Fig. 6.—Average gene frequencies of the loci of the $D$, $Q$, and $C$ subgenotypes, shown separately for the four population sizes. There are two $D$, $Q$, and $C$ subgenotypes for each run, and two runs were made at each population size.

of the gene frequencies of the $D$, $Q$, and $C$ genotypes are given in Figure 6, plotted against generation of selection for the two runs made at each of the four sizes of population.

The dominance genotypes show no consistent changes from the original frequencies except in small populations where such deviation are probably due to random sampling effects. The effect of selection on the percentage of progeny with phenotype within limits cannot be ascribed to modification of the dominance system.

The epistasis genotypes show consistent changes such that the relationship of genotype to phenotype deviates considerably from a linear function. In the runs at smaller population sizes the high incidence of genetic fixation interacts with the selection for epistasis. This is not evident at larger population sizes. There is a trend towards a gene frequency of $0 \cdot 8 \rightleftharpoons 0 \cdot 9$ in the $Q$ genotypes and of

$0 \cdot 5 \rightleftharpoons 0 \cdot 7$ in the $C$ genotypes, which correspond to a range of $-(0 \cdot 3 \rightleftharpoons 0 \cdot 4)$ for $q_i$ and of $0 \cdot 0 \rightleftharpoons -0 \cdot 5$ for $c_i$. These ranges of $q_i$ and $c_i$ correspond to a set of curves of the type shown in Figure 7.

Clearly the increased percentage of individuals with phenotypes within the selection limits has been produced by changes in frequency of the $Q$ and $C$ genotypes; these have brought about a relationship of phenotype to genotype such that the majority of $A$ genotypes have phenotypes within the selection limits of $\pm 1 \cdot 0$.
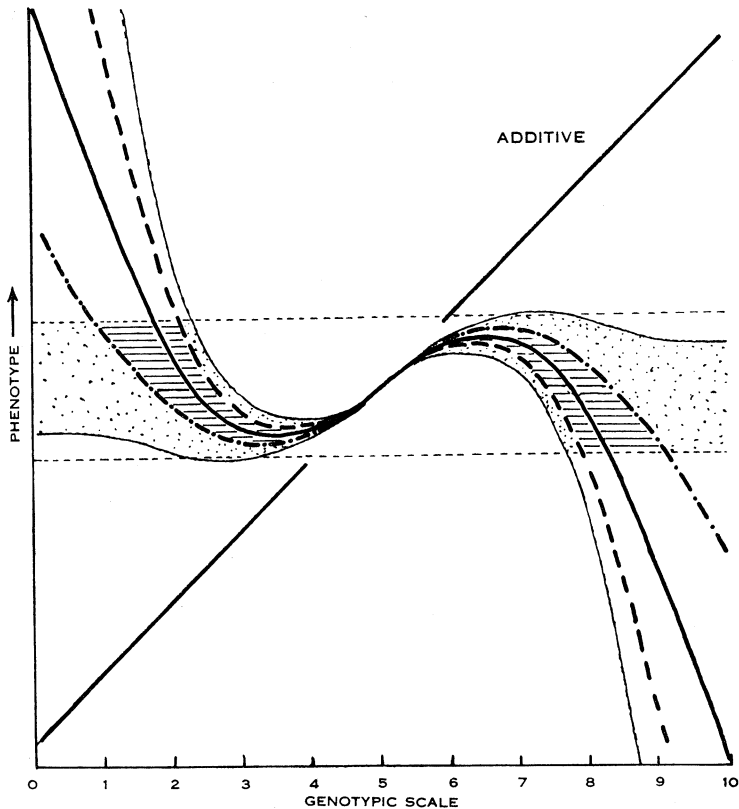


Fig. 7.—Relationship of genotype to phenotype for $q_i = 0 \cdot 4$ and $c_i = +0 \cdot 025, \ 0 \cdot 0, \ -0 \cdot 025, \ -0 \cdot 05, \ -0 \cdot 075$.

### (b) Conclusion

The runs made with this programme have shown that selection of an additive genetic system will in the absence of epistasis lead to an increased degree of genetic fixation, as shown by the straightforward mathematical analysis (see Robertson 1956). This is especially marked in runs made at small population sizes. Inclusion of a variable degree of epistasis, under genetic control, causes a marked reduction of the rate of genetic fixation for any given degree of phenotypic uniformity.

Selection of the epistasis genotypes causes a change of the genotype–phenotype relationship from the linear, additive form, to a complex form which is such that

the majority of the $A$ genotypes have the same phenotype. The effect of this on the frequency distribution of phenotype is shown in Figure 8.

Selection against extreme phenotypes can, in our genetic model, affect either the genetic system determining the genotype–phenotype relationship of the $A$ loci or the degree of fixation of the $A$ loci. It is clear that the former effect predominates. Some phenotypic variability still does occur even after 50 or more generations of selection. This is due predominantly to segregation of the $Q$ and $C$ genotypes, rather than of the $A$ genotypes. Consequently, long continued selection should, eventually, result in the fixation of the epistasis genotypes. Selection of a complex genetic system, although decreasing the rate of fixation of the additive loci, results in fixation of the epistasis-determining loci, i.e. selection against extremes results in fixation, but in a complex system this fixation is restricted to a specific part of the genotype.
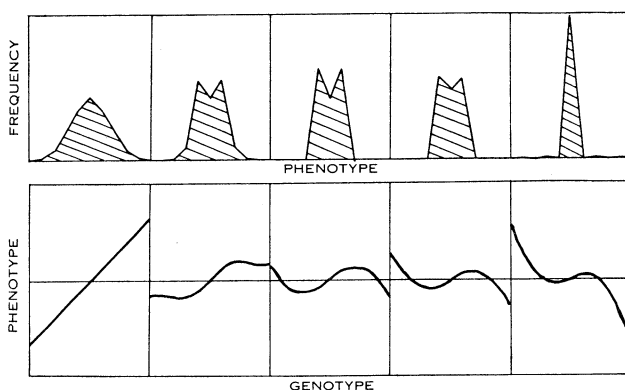


Fig. 8.—Frequency distribution of phenotypes when the frequency distribution of genotypes corresponds to an $F_2$ from completely heterozygous parents. This is shown for several genotype–phenotype relationships.

An intriguing feature of these runs has been the form of the genotype–phenotype relationship which has been produced. Such a relationship, if genetically fixed, will have unexpected effects on the progress of selection towards an extreme phenotype. Suppose that selection against extreme phenotypes has caused the fixation of such a relationship, and that the gene frequencies of the additive loci are distributed to form a normal distribution centred on a "zero" phenotype. Then selection towards either an extreme positive or an extreme negative phenotype will initially cause a straightforward shift of the distribution. However, as selection proceeds the distribution will be moved towards the inflexion point of the genotype–phenotype relationship, and this will produce a marked decrease of the phenotypic variance, due to an increase in the proportion of genotypes having the same phenotype. The phenotypic distribution will become skewed against a selection limit, and further selection for the extreme phenotype will be against the genetic extremes, resulting in fixation of the additive loci. It would seem that no further advance could be achieved. However, if the extreme genotypes of the positive-selected

population could be identified, they could act as a basis for a negative selection line in which there would be no epistasis restrictions on the effectiveness of selection.

The examination of epistatic relationships has rarely been taken past the identification of a term in the analysis of variance. This is due to the attitude that selection can only act on additive variance. If the more general view is adopted— that epistatic systems can, as has been shown for dominance systems, be affected by selection—then it is pertinent not only to determine the existence of epistasis, but also to measure the pattern of its effect. The results of Waddington (1957), Dun and Fraser (1959), and Rendel and Sheldon (1960) have demonstrated that knowledge of the pattern of epistasis is useful in the design and interpretation of artificial selection experiments. Many of the ambiguities found in selection experiments when selection pressures are reduced or reversed suggest that mechanisms of this type may operate and are certainly worth considering as a basis for experiment.

Simulation of genetic systems by programming an electronic computer is a research activity readily available to the experimenter for the small expenditure of time taken to learn the techniques of programming. As a concomitant to actual experiments it will allow the gap between the theoretical and experimental geneticists to be bridged almost effortlessly. Genetic models can be devised, programmed, and tested within weeks or months; certainly with sufficient speed for an experimenter to examine many of the theoretical consequences of his ideas before he devises experiments with live organisms. This would seem to be the field in which this method can be used to the greatest value, though its use in the methodical examination of the importance of variables determining the effectiveness of selection is undeniable. However, it is more economical to restrict such extensive studies to programmes for extremely fast machines with very large memories allowing maximum efficiency, a very necessary feature when time on a machine may total several hundred hours.

## VIII. References

DUN, R. B., and FRASER, A. S. (1959).—Selection for an invariant character, vibrissa number, in the house mouse. *Aust. J. Biol. Sci.* **12**: 506–23.

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

FRASER, A. S. (1957b).—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. (1959).—Simulation of genetic systems by automatic digital computers. V. Linkage, dominance, and epistasis. Proc. Int. Symp. Biomet. Genet. (In press.)

FRASER, A. S., and KINDRED, B. M. (1960).—Selection for an invariant character, vibrissa number, in the house mouse. II. Limits to variability. *Aust. J. Biol. Sci.* **13**: 48–58.

LERNER, M. (1958).—"The Genetic Basis of Selection." (J. Wiley & Sons, Inc.: New York.)

MATHER, K. (1943).—Polygenic inheritance and natural selection. *Biol. Rev.* **18**: 32–64.

RENDEL, J. M. (1959).—Canalization of the scute phenotype of *Drosophila. Evolution* **13**: 425–39.

RENDEL, J. M., and SHELDON, B. L. (1960).—Selection for canalization of the scute phenotype in *D. melanogaster. Aust. J. Biol. Sci.* **13**: 36–47.

ROBERTSON, A. (1956).—Effect of selection against extreme deviants. *J. Genet.* **54**: 236–48.

WADDINGTON, C. H. (1957).—"Strategy of the Genes." (Allen and Unwin: London.)