The effect of initial reverse selection upon total selection response

By J. S. ALLAN and ALAN ROBERTSON

Institute of Animal Genetics, Edinburgh, 9

1. INTRODUCTION

The effect of artificial selection on a random breeding population will be most clearly seen in a change in the mean of the character under selection, but it will also alter other properties of the population, in particular its capacity to respond to further selection of various kinds. Some loci may well have been fixed in the initial period and be unable to contribute to further change. There have been many selection experiments in which the direction of selection has been changed after some time to investigate the genetic variance remaining. There have, however, been few in which such lines have been selected continuously to the limit. During such an experiment, it was realized that we had no theoretical basis for the interpretation of the results. It therefore seemed desirable to investigate theoretically the effect on selection response in a given direction of some generations of initial selection in the reverse direction. We are then concerned with the total change produced by some generations of *reverse* selection followed by renewed *forward* selection to the limit with the same selection intensity and population size. This paper puts forward the theoretical conclusions. It is also presented as an illustration of a novel use of computers in the theoretical analysis of selection problems.

A theory of limits in continued forward selection has been developed by Robertson (1960) based on ideas put forward by Kimura (1957). In this, Wright's concept of the gene frequency distribution plays a considerable part.

For a gene with an additive effect on relative selective advantage with a difference of s between the two homozygotes, it was shown that, while the time scale of the change of the gene frequency distribution under selection is proportional to N, the population size, the pattern of this change is dependent only on Ns. This allows us to obtain general results from the computer using only one value of N and a range of values of s.

To do this, a Sirius computer was programmed to produce, first, the matrix of transition probabilities relating the gene frequency distribution in one generation to that in the next.

Each row in this matrix is a binomial distribution with index $2N + 1$ and mean equal to the initial gene frequency modified by selection. The squaring of this gives the matrix for two generations of selection and a further squaring that for four

generations and so on. By this means, we obtain the gene frequency distributions on the reversal of the direction of selection for the various initial gene frequencies.

Now we have from Kimura an explicit formula for the chance of eventual fixation of a gene under continuous forward selection, in terms of its initial gene frequency, q, as follows:

$$u(q) = \frac{1 - e^{-2Nsq}}{1 - e^{-2Ns}} \qquad (1)$$

We can then use this expression, together with the generated gene frequency distributions when forward selection is resumed, to give us the chance of eventual fixation for the whole process. This amounts to evaluating expression (1) for the $2N + 1$ possible gene frequencies and multiplying this by the matrix describing the gene frequency distributions when selection is reversed.

In the initial computer programme, which was written for $N = 5$, we produced the matrices corresponding to the second, fourth and eighth generation by squaring the original matrix three times and then multiplied each of these by the vector of the values from expression (1). The squaring of a matrix of rank m involves about m^3 arithmetic multiplications, so that in getting to the eighth generation we did about $3m^3$ such multiplications followed by a multiplication of the vector (1). It was pointed out to us that we could achieve the same end by a repeated multiplication of the vector by the matrix. Each such operation involves m^2 arithmetic multiplications, giving $8m^2$ in getting to the eighth generation. When we decided to repeat the calculations with $N = 10$, needing sixteen generations to get the same effect as eight generations with $N = 5$, we rewrote the programme in the suggested form of repeated vector multiplication. The two processes are described visually in Table 1, using A and a to represent the matrix and vector respectively.

Table 1

Generations of initial reverse selection	Method	
	(1)	(2)
1	$A \times a$	$A \times a$
2	$A^2 \times a$	$A \times Aa$
3	—	$A \times A^2 a$
4	$A^4 \times a$	$A \times A^3 a$
n	$A^n \times a$	$A \times A^{n-1} a$

2. THE EFFECT OF A CHANGE IN N ON THE COMPUTER RESULTS

To check the basic predictions that we could obtain general results from calculations made at only one value of N, we repeated some calculations with N equal to both 5 and 10. It will be recalled that the theory predicts that the pattern of change is proportional to Ns and the time scale to N. With a given value of Ns, we must compare the figures for a given number of generations of reverse selection with $N = 5$ with twice that number of generations for $N = 10$. The results corresponding to four

Table 2. *The effect of population size* N *on the computer results using* N = 5 *and* N = 10, t/N = 0·8

|	Ns = 1		Ns = 2		Ns = 4		Ns = 8	
	N = 5	N = 10	N = 5	N = 10	N = 5	N = 10	N = 5	N = 10
0·1	0·120	0·123	0·119	0·126	0·076	0·093	0·005	0·023
0·2	0·238	0·242	0·240	0·251	0·164	0·194	0·016	0·054
0·3	0·352	0·357	0·360	0·373	0·266	0·303	0·037	0·096
0·4	0·462	0·467	0·478	0·490	0·380	0·419	0·073	0·154
0·5	0·568	0·572	0·591	0·602	0·503	0·538	0·136	0·234
0·6	0·668	0·672	0·698	0·706	0·630	0·657	0·237	0·342
0·7	0·762	0·765	0·795	0·800	0·754	0·771	0·391	0·483
0·8	0·849	0·850	0·879	0·881	0·865	0·872	0·599	0·659
0·9	0·929	0·929	0·949	0·950	0·951	0·952	0·832	0·851
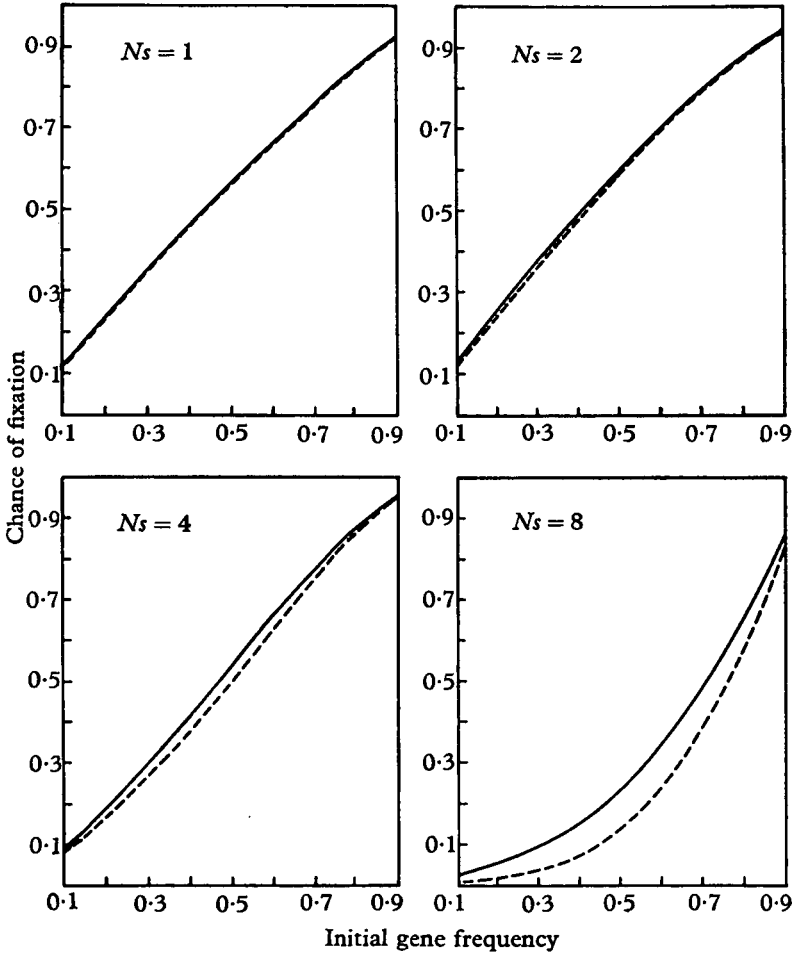
Fig. 1. The effect of population size N on the computer results,
N = 5: ······
N = 10: ———

and eight generations for $N = 5$ and $N = 10$, respectively, are given in Table 2 and Fig 1. The results are very similar for the two lower values but differ markedly at the highest value, $Ns = 8$, for which $s = 1 \cdot 6$ when $N = 5$. The method must obviously break down as s increases beyond 2 as we will be then faced with some apparent negative gene frequencies after selection. In the subsequent discussion, we shall use the results calculated for $N = 10$.

3. THE THEORETICAL EXPECTATIONS FOR SMALL VALUES OF Ns

In the earlier paper on selection limits, explicit expressions for the chance of fixation at the limit were derived for the situation in which Ns, and consequently the mean change in gene frequency, was small. We can use the same treatment here. We then proceed on the assumption that because of the restricted population size the average value of $q(1-q)$ declines by a proportion $1/2N$ each generation so that after t generations, it will have the value of $q(1-q)e^{-t/2N}$. Under continuous forward selection, the total expected change in gene frequency is then $2N$ times the change in the first generation. After t generations of reverse selection, we would expect the mean gene frequency to have been reduced by $Nq(1-q)(1-e^{-t/2N})$ while the mean value of $q(1-q)/2$ has been altered by a factor of $e^{-t/2N}$. The subsequent change on the resumption of forward selection will then be $2N$ times this latter value, giving for the total change in gene frequency

$$Nq(1-q)\left(2e^{-t/2N} - 1\right)$$

Putting $t = 0$ in this expression, implying no reverse selection, we find the expected expression $Nq(1-q)$ for the effect of continuous forward selection. If $t = \infty$, implying continuous reverse selection, we obtain the same expression with a negative sign, as expected. We see that the overall chance of fixation decreases as t increases and will be equal to zero when $e^{-t/2N}$ is equal to one-half. This corresponds to $t = 1 \cdot 4N$ generations. Thus, if the reverse selection lasts for longer than this, we will not get back to the initial starting point. In the earlier paper it was shown that the expected half life of a selection process (the number of generations for half the total gene frequency change to be produced) had this same value, if Ns is small. We might expect that, for small gene frequency changes, the half-life for continued forward selection would then be the same as the point of no return under reverse selection.

4. THE COMPUTER RESULTS

Figure 2 shows the chance of fixation at the limit for various periods of reverse selection. The theory presented above gives us a prediction of the slope of these curves when Ns is small. This should then change sign at $t = 1 \cdot 4N$. For all initial gene frequencies, the curves do have a positive initial slope for the four smaller values of t but negative slopes when $t = 1 \cdot 5N$. It will be seen that the effect of the reverse selection is greater for desirable genes with low initial frequencies and for

those with large effects. The first point is in agreement with the effect of the restriction of population size without selection discussed in the earlier paper. A gene at a low initial frequency in the population has a high probability of being lost entirely after some generations of restriction in population size. Those desirable genes with higher frequencies but with large effects will have their frequencies rapidly reduced by the reverse selection until they are brought into the region in which the chance of a complete loss from the population is high.
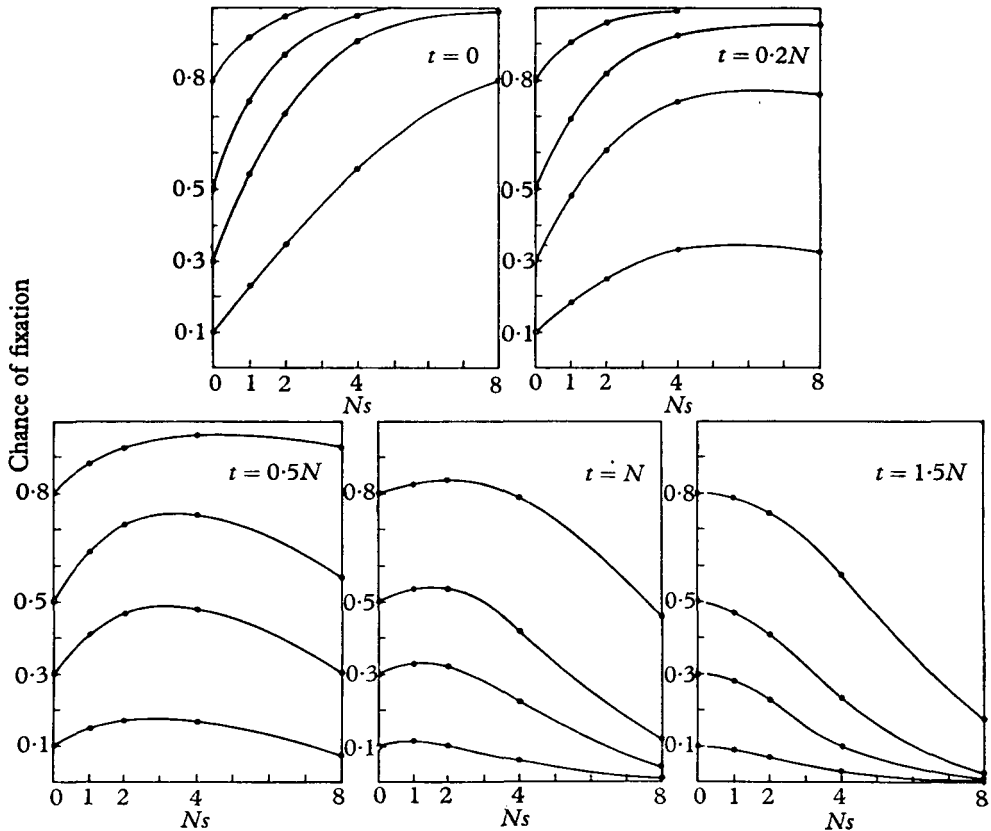

Fig. 2. The chance of fixation for various initial gene frequencies, values of Ns, and durations of initial reverse selection t.

We discussed earlier the 'point of no return'—the number of generations of reverse selection which made it impossible for resumed forward selection of the same intensity in a population of the same size to return the mean gene frequency to the starting point. It is obvious from Fig. 2 that the value of $1.4N$ generations deduced for small values of Ns is an upper limit and that genes with larger effects reach the point of no return rather sooner than this. Figure 3 shows the point of no return plotted against initial gene frequency for different values of Ns. This has an obvious similarity to Fig. 3 in the earlier paper which showed the effect of Ns on the half-life of the selection process.

One interesting and unexpected point emerges from Fig. 2. The curves for $t = 0.5N$ show, for instance, that the chance of fixation plotted against Ns passes through a maximum when Ns lies in the neighbourhood of 3 to 4 for all initial gene frequencies and then declines. When $t = N$, these maxima occur at lower values of Ns between 1 and 2 and for the highest value of t, there are, of course, no maxima as all curves have a negative slope at $Ns = 0$.
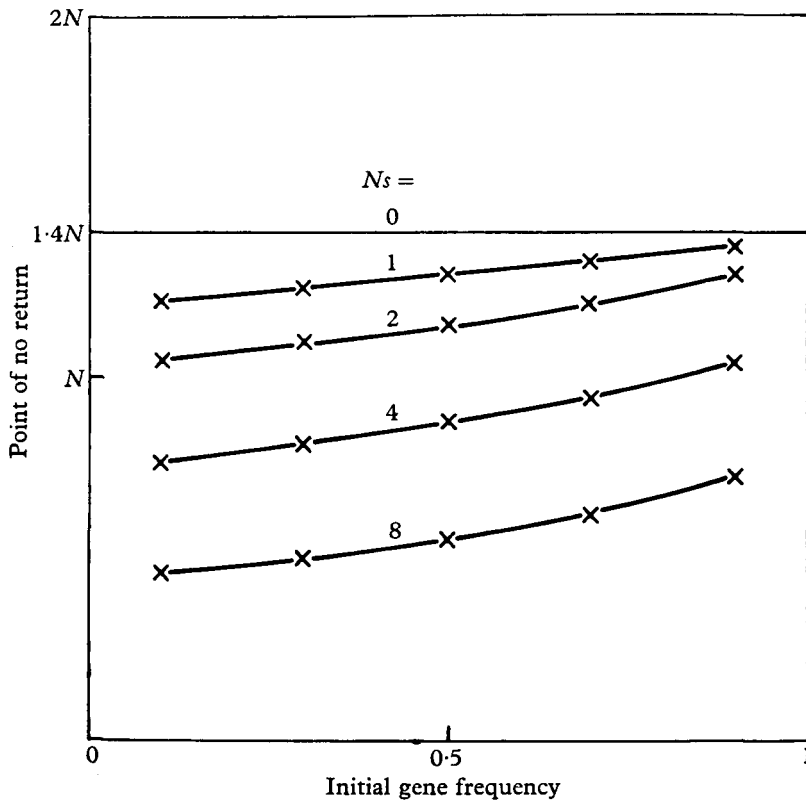

Fig. 3. The duration of reverse selection, which makes it impossible to return to the initial frequency using the same population size and selection intensity.

5. THE RELATIVE REDUCTION IN TOTAL ADVANCE PRODUCED BY THE REVERSE SELECTION

We have so far been dealing entirely with gene frequencies. In selection experiments on continuous characters, these are generally not observable and we can merely measure changes in the mean of the selected character. How can we apply our present results to continuous variation? Under artificial selection, the effect of a gene on the selected character is proportional to its consequent selective advantage, so we may treat genes having high s values as being those with large effects on the selected character. The charges in the mean of the selected character then reflect changes in the frequencies of the genes affecting it, each weighted by the

magnitude of the gene effect. We must therefore examine our curves again in terms of the relative reduction in the advance under selection caused by a given number of generations of reverse selection which is shown in Fig. 4. The horizontal line in each graph gives the theoretical expectation for small values of Ns. When $t = 0 \cdot 2N$, the loss in advance is highest either at low frequencies with high values of Ns or at high
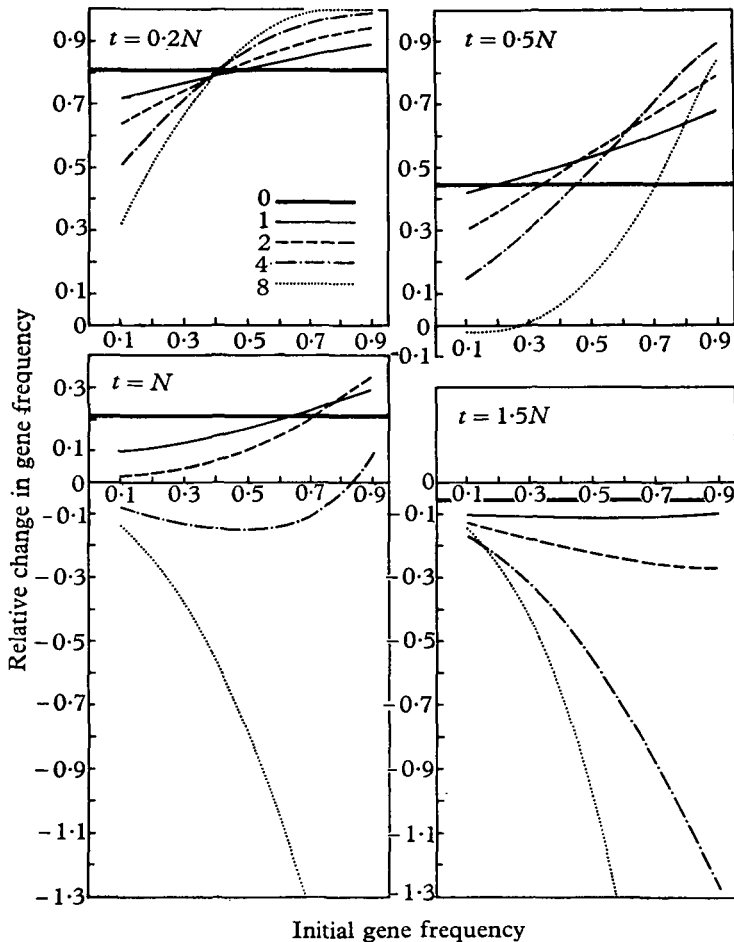

Fig. 4. The expected change in gene frequency, expressed as a proportion of the change under continuous forward selection (derived from Fig. 2). The curves refer to different values of Ns.

gene frequencies with low values of Ns. If we therefore average our effects over gene frequencies assuming a rectangular distribution, we would find that the reduction was almost independent of Ns. But for longer periods of reverse selection, Ns becomes more important. In the two lower graphs, we see that for the higher values of Ns the curves decline very steeply at high gene frequencies. For short periods of reverse selection the reduction in advance will come mostly from genes of

low initial frequency but, as the period of reverse selection becomes of the order of N itself, the loss will come from genes of large effects irrespective of their frequency.

6. THE SHAPE OF THE RESPONSE CURVES

We have discussed so far only the total response to selection but response curves themselves are of some interest. If we resume forward selection after a given number of generations of reverse selection, for instance, will the resultant population gain on or lag behind that which has been selected continuously in the forward direction. We can easily calculate the mean gene frequency when reverse selection is stopped and we have of course already obtained values for the gene frequencies at the limit.
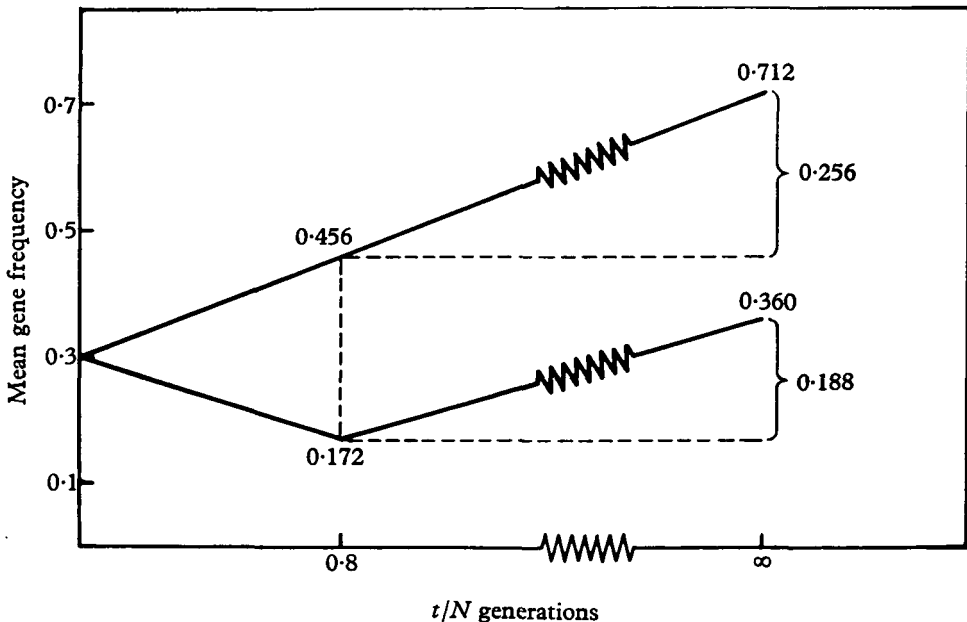

Fig. 5. An example of the divergence of lines after renewed forward selection.

One such situation is presented schematically in Fig. 5 for $Ns = 2$, $t = 0.8N$ and an initial gene frequency of 0.3.

We find that the results could be generalized as follows. If the desirable gene was at an initial gene frequency greater than 0.5, then the line, on resumption of forward selection, converges on the line which has been selected forward continuously. If, the initial gene frequency of the desirable gene was less than a half then the lines diverge on the resumption of forward selection. This is an obvious consequence of the fact that $q(1-q)$ is at a maximum when q is equal to 0.5. If the initial value of q is greater than this, then the line selected forward continuously is moving into regions of lower values of $q(1-q)$, whereas the line selected in reverse is moving to higher values. As a result, when forward selection is resumed, the genetic variation in the reverse line will be greater than that in the forward line and the two will converge. Similar arguments apply for desirable genes at low initial frequencies.

A more detailed and mathematically more sophisticated approach was used in one particular instance by using the computer to generate the actual response curves. The matrix corresponding to any number of generations of selection represents the series of gene frequency distributions deriving from each possible initial gene frequency. When we multiplied this by the vector (1), we then obtained the chance of fixation at the limit, but if we multiply it instead by the vector of gene frequencies, i.e. for $N = 5$, the vector 0, 0·1, 0·2 and so on, we obtain the instantaneous gene
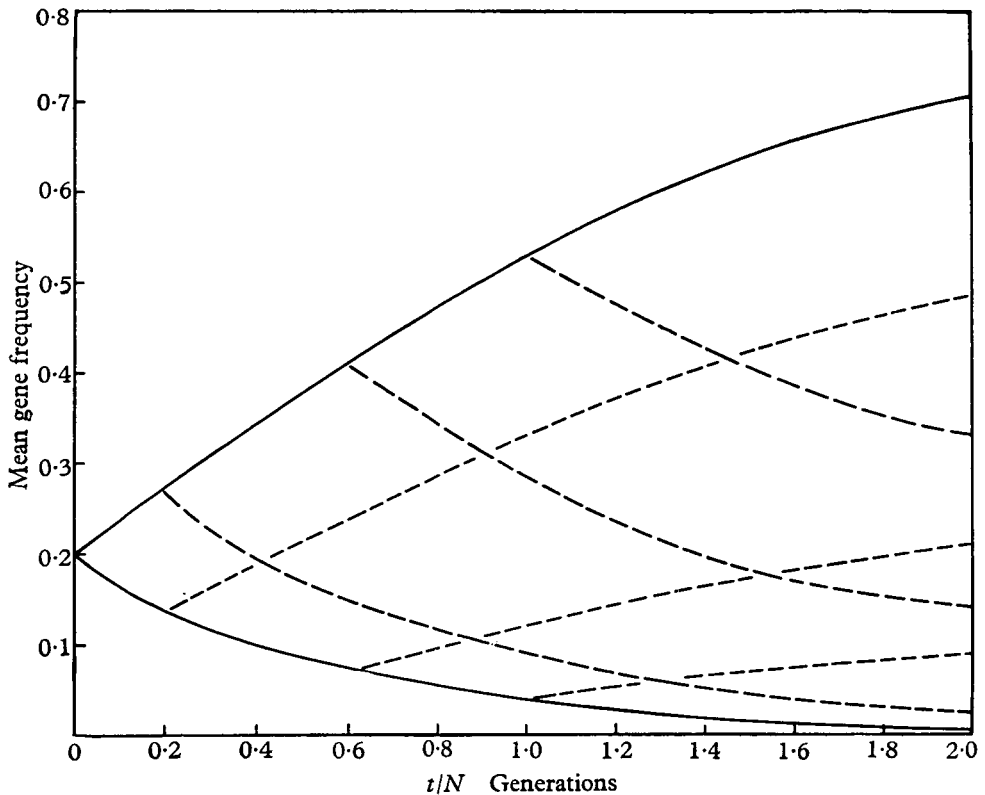

Fig. 6. Selection response curves for $Ns = 4$ and various periods of reverse selection, calculated by the computer.

frequency in the selected line. This calculation was done by modifying the initial programme and substituting the new vector for the vector (1). We did this only for Ns equal to 4. We were than able to generate the curve of the change of mean gene frequency with time for any initial gene frequency and with any number of genera- tions of reverse selection. The results for an initial gene frequency of 0·2 are given in Fig. 6. The two solid lines are the curves for continuous selection. The dotted lines represent the changes in gene frequency when selection is reversed after $0·2N$, $0·6N$ and N generations respectively. The figure illustrates very well the general points made in the preceding paragraph. Although the long-term response curves are, of

course, asymmetrical, it should be noted that the immediate response to selection in the two directions are symmetrical in all cases.

In experiments involving an initial period of reverse selection, the lines have frequently only been selected forward again for a comparatively few generations and there are no examples in the literature in which the resumed forward selection has reached the limit. As a consequence this tendency to approximate parallelism of subsequent response has not received much, if any, attention in the past. This theoretical result has in fact been confirmed in the experimental part of this investigation which will be published later.

This analysis has been restricted to independently segregating genes with additive action. We can to some extent remove the latter restriction by applying the 'half-life' results for low Ns values in the earlier paper to the point of no return. We can then say that, as an upper limit, the point of no return will be reached in $2 \cdot 12N$ generations for a rare recessive and in $1 \cdot 03N$ generations for a frequent one. The computer programmes could be modified to deal with recessive genes, merely by putting $sq^2(1-q)$ for $s/2\, q(1-q)$ as the mean change in gene frequency produced by selection.

7. DISCUSSION

The use of computers in the theoretical analysis of selection problems is of course not new, but, in other approaches, the use has been at a direct rather than at an abstract level. In applying the computer to any problem, it is obviously best to get as far as possible by algebra and to resort to the computer only when this fails. We may therefore use a computer in genetic simulation work at several different levels. The most direct level, in which the digit of the computer actually simulates a gene, and in which 'organisms' are produced, replicated, subjected to environmental variation and selected, is represented by the papers of Fraser (1957) and Cockerham & Martin (1960). In this approach the digital computer is being used almost as an analogue computer, with the digits as genes. With some loss of generality, it is possible to deal with problems involving linkage in artificial selection without actually forming the diploid organisms but by applying the selection pressure at the gamete level. This is possible only for genes showing additive action and programmes have so far been written only for three linked loci although an extension would be simple. In this present paper we have used an even more abstract approach in dealing with a fairly simple problem in that we use the computer merely to derive the gene frequency distributions.

The different approaches obviously vary in the generality of the questions asked, the time it takes to get an answer, and in the relevance of the answers that are obtained. The first method is obviously extremely direct but may be time-consuming in that it may simulate processes which, in that particular problem, could well be abstracted by an algebraic expression. This is done in the second of the programmes sketched here but with an obvious loss of generality. The present method, representing an even higher degree of abstraction, is again more restricted

but has one tremendous advantage over the other two. They deal with random processes by introducing random elements into each set of results. The results are therefore subject to stochastic variation and replicates have to be run for a full understanding of the situation. But in the present approach, in which the randomness is dealt with algebraically, we derive the whole gene frequency distribution and get both mean and variability in one operation. The method can easily be extended to deal with different forms of gene action, though not of course with epistatic action, and a repetition on a much larger scale of work such as that of Zucker (1960) would be very simple.

Perhaps the most interesting aspect of these results is the demonstration that very few generations of selection can alter considerably the properties of a population—in particular its response to selection in the opposite direction. This is most clearly seen in Fig. 6, in which as little as $0.2N$ generations of selection in one direction make it impossible to attain the original limits to selection response in the other. If the initial period of selection approaches in generations the population size itself, it may be difficult to return to the original population mean. This 'point of no return' is reached for genes with small additive effect at $1.4N$ generations and the time necessary decreases as the gene effect increases.

The simplest explanation of these results derives from the high probability of loss from the population of genes at low frequency. For short durations of reverse selection, the loss will be mainly dependent on initial frequency and not on gene effect, but as the initial period of selection increases, we begin to lose genes whose frequencies were not originally in this region but which have been drawn into it by selection. This loss is then greater for genes with large effects.

SUMMARY

A computer has been used to investigate the effect of an initial period of reverse selection on the subsequent response of a population to renewed forward selection with the same population size and selection intensity. As the computer was used to derive gene frequency distributions, there was no random element in the results obtained. A theoretical solution to the problem was obtained for genes with small effects.

The process can be adequately described by the duration of the reverse selection (expressed in terms of the population size N), the product of population size and gene effect, Ns, and the initial gene frequency. If the duration of reverse selection, t, is less than $N/2$, the loss in selection advance due to the reverse selection is roughly t/N, though slightly greater than this for genes with low frequency. The 'point of no return' after which it is impossible, with the same population size and selection intensity, to return even to the starting frequency is $1.4N$ generations for genes with small effect and this declines as the gene effect increases.

Some extension of results to recessive genes is also given.

REFERENCES

COCKERHAM, C. C. & MARTIN, F. G. (1960). The I.B.M. 650 and its use in selective studies. *Biometrical Genetics.* Pergamon Press.

FRASER, A. S. (1957). The simulation of genetic systems by automatic digital computers. II. The effects of linkage on the rates of advance under selection. *Aust. J. Biol. Sci,* **10,** 492–99.

KIMURA, M. (1957). Some problems of stochastic processes in genetics. *Ann. math. Statist.* **28,** 882–901.

ROBERTSON, A. (1960). A theory of limits in artificial selection. *Proc. roy. Soc.* B, 234–249.

ZUCKER, LOIS M. (1960). A model breeding experiment for two-way selection in small populations showing directional dominance. *Genetics,* **45,** 485–498.