SELECTION THEORY VERSUS SELECTION RESULTS - A COMPARISON

J.W. James
Department of Wool and Animal Science, University of New South Wales,
Kensington, N.S.W., 2033, Australia

SUMMARY

INTRODUCTION

Selection experiments are carried out for many reasons. In some instances, the aim may be to develop genetically different lines for other purposes, such as physiological research, the progress of selection being of no inherent interest. In other cases, a selection experiment may be intended to help elucidate the genetical control of a particular character. A good deal of selection in mice is carried out for this motive, where the mouse is regarded as a model for the pig or other mammals, or is of interest for itself (McCarthy, 1980). However, a number of experiments have been conducted to evaluate quantitative genetic theory. Such experiments are particularly difficult, because the comparison of experimental results with theoretical predictions requires substantial replication to be used if the results are to be informative, since both prediction and empirical result are liable to errors of estimation.

The paper by Clayton et al. (1957) set a new standard for such experiments, in the thoroughness of the comparison between theory and observation and in the replication of selection lines. In essence, the approach was to estimate correlations between relatives in the base population and check that these showed a pattern consistent with quantitative genetic theory ("static model"). Then, since the correlations were consistent with theory, the estimate of heritability was used together with simple selection theory to predict response to individual selection of varying intensity and to family selection, over a period of five generations ("dynamic model"). It was recognised that selection would modify the properties of the population, so that it was necessary "...to make a subjective judgement as to whether we shall consider any discrepancies important or not." Some factors which affect continued application of theory, such as changes in gene frequency, remain intractable, but some processes, such as generation of gametic unbalance by selection (Bulmer, 1971) and effect of selection on rate of inbreeding (Wray, 1989) are now better understood, and perhaps could be included in predictions so as to reduce discrepancies to which subjective judgement must be applied.

Sheridan (1988) has reviewed experiments in both laboratory and farm animals in which preliminary and realised heritability estimates were available. Of 198 comparisons, 54% differed more than 30% of the realised heritability, and 38% differed by more than 50%. Realised heritabilities differed substantially with direction of selection in about half of the bidirectional experiments. Approximate tests of significance at the 5% level indicated that in 25% of cases differences between preliminary and

realised estimates were significant. It is not clear that standard errors have been correctly calculated, allowing for genetic drift, in all cases, so the number of statistically significant discrepancies may have been overestimated. Sheridan concluded that there was "...a general lack of good agreement between estimated and realised genetic parameters." Here, "estimated" means preliminary. One important conclusion which may be drawn from Sheridan's survey is that inaccurate estimates of base population heritability make poor predictors of response, while inaccurate estimates of realised heritability give poor tests of prediction. Realised heritability was less than predicted in 57% and greater than predicted in 38% of cases. Since most known factors would tend to make realised heritability less than predicted, a major cause of discrepancy would appear to be sampling error. Falconer (1973) in a replicated selection experiment for six week weight in mice made a preliminary estimate of within-litter heritability with 281 degrees of freedom for sires, and found a value of 0.113 with a standard error of 0.110. He also estimated the within-litter heritability from offspring-parent regression as 0.468 with a standard error of 0.063. These two very different estimates from the same collection of data would usually cause great uncertainty in making predictions. A separate analysis on the same strain (Monteiro and Falconer, 1966) gave sire component heritability estimates very similar to the regression estimates of Falconer (1973), which would therefore be favoured. In the event, the average realised heritability was 0.37 over ten generations, in good agreement with the regression estimate. But in the absence of the other data, the value to use for prediction would have been unclear. Thus, while sampling error may explain much of the discrepancy, there are many examples in which agreement is poor. It is therefore worth considering whether more recent theoretical developments can account for a large fraction of the disagreements observed.

REDUCED SELECTION DIFFERENTIAL

It has been widely appreciated for a long time that in small populations the selection differential for a given fraction selected is smaller than in a large population, and suitable tables based on normal order statistics have been provided. More recently, Hill (1976) and Rawlings (1976) have shown that family structure will lead to further reduction in selection differential, the effect being stronger as the correlation between relatives increases, and this effect is now well recognised. It is a little less well recognised that with individual selection the genetic response is further reduced "...because the best individuals on phenotype in the population are best in their own family, so the simple regression on phenotype overestimates family effect" (Hill, 1977). We can illustrate the magnitude of these effects in a couple of simple cases where 64 animals are measured, consisting of either 16 families of 4 of which 1/4 progeny are selected, or 8 families of 8 from which 1/8 progeny are selected. These are assumed to be full sib families, with heritabilities of either 0.2 or 0.5, and no common environment effects, so that the correlations between sibs are either 0.1 or 0.25. Using Table 1 of Hill (1977) one finds:

Table 1 Fraction of Infinite Population Response

Families	H2	Finite SD	Family SD	Full Effect
8	0.2	.9803	.9745	.9250
	0.5	.9803	.9643	.9275
16	0.2	.9857	.8835	.9638
	0.5	.9857	.9794	.9660

Finite SD: taking account of finite numbers on the selection differential.
Family SD: adding the effect of family structure on selection differential.
Full effect: adding effect of family genotypic value.

Notice that the effect of the final reduction is less with higher h^2. Actually, the correction to family effect is larger with higher h^2, but the contribution of family effect to breeding value is less when h^2 is high. With index selection using both individual value and family average there is no correction to family effect, so only the Family SD needs to be considered. However, the correlation between relatives is higher for the index than for individual selection, so there is a greater reduction in selection differential. This theory is relevant to the interpretation of experiments comparing individual selection with index selection such as those of Wilson (1974), but existing tables are not extensive enough to allow a proper comparison to be made. However, in the above table it can be seen that the full effect with a low correlation between relatives may be greater than the Family SD effect with a high correlation, and it is possible that the effect on individual selection may be greater than that on index selection, though the selection differential would be more affected for the index.

In the experiments of Wilson, the first series were conducted over several generations, but the second series were done for only a single generation, in order to avoid the problem of greater inbreeding with index selection than with individual selection. In the single generation experiments, the average responses in ten replicates for each of larval weight and pupal weight were slightly, but not significantly, greater for individual selection than for index selection (pupal weight 11.59 vs. 10.77; larval weight 11.46 vs. 10.98). Given an expected superiority of about 20% for index selection using simple theory, the magnitude of the effects discussed by Hill is not sufficient to account for the full difference, though it may contribute something. The actual expected superiority of index selection would also be a little less than calculated because the calculated value assumes that correlations between relatives are known exactly instead of being only estimated, but it is well known that the error so introduced is not very important. A good explanation for the apparent inefficiency of index selection, apart from sampling error, is not easy to find.

A similar, as yet unpublished, experiment was done in my own laboratory using sternopleural bristle number of Drosophila melanogaster. With full sib families of size ten and a correlation between sibs of 0.17, the

expected index efficiency was 21% greater than individual selection. Averaged over 29 single-generation selection lines (one of 30 replicates was lost), the observed responses were 0.93 bristles for individual and 1.24 bristles for index selection. The ratio of 1.33 with standard error 0.097 does not differ significantly from expectation, but is highly significantly different from 1.0. Thus in this case, unlike that of Wilson, the agreement between theory and observation is rather good, with index selection clearly better than individual selection, and there is no need to try to improve the fit by considering more complex theory, which would not much affect the conclusions.

GAMETIC UNBALANCE AND INBREEDING

The generation of negative correlations between breeding values at different loci by selection, the so-called "Bulmer effect", is expected to reduce the effective heritability in lines subjected to directional selection, with the magnitude of the effect being greater for more intense selection. It is thus somewhat surprising that in the experiments of Clayton et al. (1957) the ratio of observed change to predicted change declined steadily as selection became less intense, both for up and for down selection. The discrepancy was especially marked for selection of 80%. The same number of parents was selected at each intensity, so effective population size would be expected to be somewhat larger with less intense selection (Robertson, 1961). Thus both the expected effects of gametic unbalance and of inbreeding are in contrast to the observed trend.

On the other hand, in the work of Frankham et al. (1968a) the realised h^2 values, averaged over parental numbers, were 0.159, 0.156 and 0.176 for 10%, 20% and 40% selected, but only 0.014 for 80% selected. Thus in both experiments the realised response was particularly poor for weak selection, and resulted in discontinuation of the 80% lines at generation 12 by Frankham et al. Perhaps because the artificial selection pressure applied is so weak in this case, a small opposing force of natural selection may be able to exert a proportionately greater influence on response than when selection is more intense. Certainly, the results of relaxed selection by Clayton et al. suggest only a weak natural selection towards the original mean. However, the Frankham et al. results do not show the consistent trend found by Clayton et al. One possible reason is the time period involved, 5 generations for Clayton et al. and 12 for Frankham et al., so that inbreeding effects had more time to operate in the second experiment, but the graph of Frankham et al. does not suggest that a very different picture was seen at 5 and 12 generations. The apparent ineffectiveness of weak selection perhaps merits further study, though such study is unattractive because of the slow responses to be expected, and the large standard errors relative to the small responses.

Since Frankham et al. gave realised h^2 values for different population sizes and selection intensities, their results are particularly suitable for investigating the effects of inbreeding and gametic unbalance on response. Following the analysis of Sheridan et al. (1968), they used

h^2=0.15 for prediction, but I have preferred to use h^2=0.25, based on the sire and dam component estimates given by Frankham et al. (1968b), as this appears more reasonable in the light of their results.

Let us first consider the inbreeding effects. Robertson (1961) showed that inbreeding would be greater with more intense selection, but it was soon recognised that his simplified treatment overestimated the effect. More recently, Wray (1989) has reported a more detailed analysis and shown that inbreeding in computer simulations can be well predicted using her methods. It is, of course, always possible to get predictions by computer simulation, but less time-consuming methods are desirable and often more instructive. Here I shall use an approximate formula taken from Wray's work,

$$2N\ \Delta F = 1 + Q^2 i^2 h^2\ (1 - kh^2)/(2 - kh^4)$$

where i is the standardised selection differential, k is the proportional reduction in phenotypic variance by selection, and Q^2 is a measure of the spread of genes in descendants. For the range of selection intensities in the Frankham et al. experiments, Table 6.5 of Wray (1989) suggests that a value of Q^2 = 3 would be approximately correct, though a value varying with selection intensity would be more accurate. Using h^2 = 0.25 and values of i and k from normal distribution theory gives 2N ΔF as 1.9397 for 10%, 1.6061 for 20% and 1.2961 for 40%. In the absence of selection there would be some reduction in ΔF because parents were chosen from different bottles, and some increase because of variation of family size within bottles. It is difficult to quantify these effects, but I have arbitrarily increased the ΔF above by a fraction 0.1 to account for them. It could be a useful exercise to estimate the relation between N and ΔF due to these factors by analysis of variation between control lines, though more lines than are usual would be needed to get an accurate estimate. The average responses over 12 generations would be roughly proportional to $[1 - (1 - \Delta F)^{12}]/12\Delta F$, giving the values:

Pairs of Parents	Percentage Selected		
	10	20	40
10	.7532	.7894	.8257
20	.8655	.8870	.9077
40	.9300	.9417	.9525

These results allow only for inbreeding. We should also allow for gametic unbalance using the methods of Bulmer (1974) where allowance for linkage is included. Unlinked loci quickly reach equilibrium, but linked pairs do so much more slowly, and eventually make greater contributions to the negative correlation between loci. The combined effect is proportional to the harmonic mean of recombination fractions, which Bulmer calculated as 0.1 for Drosophila melanogaster. We can use this value to find the steady state gametic unbalance, but this will be too large for early generations, so I have arbitrarily averaged the expected initial and final heritability

values, which would give in an infinite population heritabilities of 0.2075, 0.2088 and 0.2114 for 10%, 20% and 40% selected. Ignoring interaction between inbreeding and gametic unbalance, the expected realised heritabilities have been calculated by multiplying the two sets of figures, with the following predictions of realised heritabilities.

Pairs of Parents	Percentage Selected		
	10	20	40
	PREDICTED		
10	.1563	.1648	.1746
20	.1796	.1852	.1919
40	.1930	.1966	.2014
	OBSERVED		
10	.1548	.1306	.1523
20	.1343	.1799	.1710
40	.1872	.1584	.2032

The agreement is not startlingly good, and there appears to be a tendency to predict too high values, perhaps because the original estimate chosen is too large. Inspection of the table may also suggest that population size effects are greater than predicted, but given the variation between replicate lines perhaps no better agreement could have been expected. The weighted mean square error of prediction of group mean h^2s is 0.0019, while the pooled within treatment mean square is 0.0014. These values, with 9 and 18 degrees of freedom, are reasonably similar, though pooling of variance estimates with differing expectations is of course involved.

FOUNDER EFFECTS AND MUTATIONS

The work of Frankham et al. (1968a) and Jones et al. (1968) was designed as an experimental evaluation of the main predictions of Robertson (1960). Another experiment done by D.E. Robertson (1969) was designed to study the effects of number of founders as outlined in this theory and extended by James (1971). Selection was for left side sternopleural bristles in D. melanogaster, and lines were begun with founder samples of 1, 5 or 20 pairs of flies from the Canberra base population. The progeny of each founder sample were randomly divided, one part being selected at once, the other after 5 generations of random mating to allow recombination. This lag had no effect on responses, so the two sets are considered here as simple replicates. A summary of the results given by James (1971) was as follows:

Numbers selected in each sex	Founder sample size		
	2	10	40
5/25	2.6 (16)	4.6 (12)	4.0 (12)
20/100	7.2 (12)	7.2 (8)	6.3 (8)

The number of lines per treatment is given in parentheses, and the values in the table are average gains in bristle number after 30 generations of selection.

The simplest theoretical prediction is that the limit to selection response is proportional to (1-1/2n) where n is the founder sample size. Although the 5/25 lines are in reasonable agreement with this theory, the 20/100 lines show no significant effect of founder sample size. These observations were interpreted as showing that gains in the 5/25 lines had been dependent on fixation of genetic variants present in the founder samples, but that response in the 20/100 lines had been very strongly influenced by selective use of variation arising during the selection process. Some simple calculations suggested that not too many "mutational" events would be required to explain the results, but no attempt at a quantitative prediction of mutational effects was made. However, following the work of Hill (1982) it seems worthwhile to check some simple predictions for these results.

If V_A is the additive genetic variance in one generation and V_M is the new additive variance arising from mutation each generation, the additive variance in the next generation is $(1 - 1/2N)V_A + V_M$, where N is the effective population size, so after t generations the additive genetic variance will be

$$(1 - 1/2N)^t V_A + 2NV_M[1 - (1 - 1/2N)^t].$$

If response in any one generation is $iV_A/\sigma P$, where i is the standardised selection differential and σP the phenotypic standard deviation, the cumulative response over t generations is

$$(i/\sigma P)[2NV_A(1 - (1/2N)^t) + 2NV_M(t - 2N(1-(1-(1/2N)^t)))].$$

For a line with n founders, V_A in this expression must be multiplied by (1 - 1/2n). Therefore the response over t generations can be written as

$$2N(iVA/\sigma P)[(1-1/2n)(1-(1-1/2N)^t)+(V_M/V_A)(t-2N(1-(1-1/2N)^t))].$$

This derivation is based on the assumptions that genes have small effects on the selected trait, and that mutations do not affect fitness. As t increases, the total response approaches

$$2N(iV_A/\sigma P)[1 - 1/2n + (V_M/V_A)(t - 2N)]$$

and in the long run the effect of founder number declines.

From calculations made in connection with the experiment of Frankham et al. (1968a), the effective population size during selection should be about 0.6 actual size in this case, so that 2N = 12 for the 5/25 lines and 48 for the 20/100 lines. Dividing by $iV_A/\sigma P$, the response in the first generation,

and setting t = 30 in the expression for total response, we find the following results.

Number selected in each sex	Number of founders		
	2	10	40
5/25	8.34 + A	10.56 + A	10.98 + A [A = $227V_A/V_M$]
20/1000	16.86 + B	21.35 + B	22.20 + B [B = $361V_M/V_A$]

Lynch (1988) reviewed work on estimation of mutational variance, and a typical value for bristles in D. melanogaster would be $V_M = 0.001V_E$ where V_E is the environmental variance. Since h^2 for one-side sternopleural bristle number in the Canberra population was about 0.2, V_M/V_A is about 0.004. Assuming an initial selection response of about 0.25, since V_P = 1.0 approx., we get the following predicted values at generation 30.

Number selected in each sex	Number of founders		
	2	10	40
5/25	2.99	3.55	3.65
20/100	5.66	6.78	6.99

These figures could clearly be modified by varying our assumptions about N or V_M/V_A, neither of which is known with precision, especially the latter, but with such untested assumptions it is not profitable to try to adjust the estimates in order to find closer predictions. Even the simple attempt to make allowance for mutation has improved agreement with experimental results, but the data show a larger effect of founder number in the 5/25 lines and a smaller one in the 20/100 lines than predicted, though in view of the variation between replicate lines it is best not to be too dogmatic about this.

No attempt to include the gametic unbalance was made either. One assumption that was not met by the data was that genes had small effects, since some genes of large effect appeared and were quickly fixed. If this were properly allowed for, perhaps agreement might improve further.

It is likely that at least some of the mutations that occurred were not "classical point mutations". In one sample of 10 founders selected at 20/100 in each sex, both the directly selected and the lag lines showed a sudden sharp response to selection after some 10 to 15 generations of selection. The pattern of extra bristles in these lines had a characteristic appearance, not present in any other line, and when the lines were later crossed and selection practised on the cross, there was no addition of responses attained by the two lines. It seems clear that the same variant was fixed in both lines, and in view of its large effect, it was not segregating in the founder sample. Since there were 34 separate founder samples, it is not plausible to attribute to chance the occurrence of the same variant in both lines from the same founders. The occurrence of this mutant would then seem to depend on rearrangement of genetic material present in the founder sample. One possibility was recombination

of genes linked in repulsion, and another, which we then favoured, was intragenic recombination. However, there are other mechanisms then unknown to us, such as transposon-induced variation. Since the occurrence of mutations may depend on the genetic material with which an experiment is conducted, the use of a single highly inbred line as a base population for evaluating the contribution of spontaneous mutation to selection response in an outbred base population may be misleading if interpreted too simply. Such lines may, however, be useful for estimating mutational variance arising from classical causes or in lines selected over an extremely long period, such as those of Yoo (1980), when genetic variation originally present will have disappeared. The use of several founder inbred lines in such experiments may be wise, as a check on whether the amount of mutational variance depends on the original genetic material.

DISCUSSION

In this survey, I have made no attempt to be comprehensive, but have rather chosen a small number of experiments to illustrate the way in which recent theoretical developments could be used to make more complicated predictions than the relatively simple ones used at the time. The predictions made here are still oversimplified for two reasons. The first one is that only the so-called infinitesimal model can be used for these predictions, since for other models data on gene frequencies and gene effects, and perhaps linkage associations, are needed and such data are simply not available. Therefore if precise predictions are to be made they must be based on the infinitesimal model. This is not to say that predictions cannot be made from other models, simply that the parameters required cannot be estimated with current methodology. The second reason is that the actual predictions have been made approximately from the chosen models, rather than with a completely thorough analysis. This has been done because I think that it is often more instructive to see an approximation which is understandable than a very complex calculation whose meaning never becomes clear. In the present context, it is always possible to run a computer simulation of a selection experiment in which the genetic model is chosen and the experimental protocol is applied to that model. It is likely, however, that such a procedure would not reveal clearly which aspect of the assumptions has caused any particular observed pattern. A large amount of simulation should enable this problem to be overcome, and it may well be that the great commitment of time which is needed to conduct a well-designed trial in quantitative genetics can justify a detailed preliminary simulation. Indeed, for some purposes a simulation experiment is superior to an experiment with real organisms. But in the end we want our theory to apply to real animals, and no matter how well it works in our simulations it is only satisfactory if we can apply it to animals.

To what extent can we say that the more highly developed theory has helped in the interpretation of the experiments reviewed? It has not helped at all in explaining the ineffectiveness of weak selection, but has been useful, in my opinion, in explaining some of the features of responses in experiments using different selection intensities and population sizes, and

also in allowing for mutation, though the improvement has not been very remarkable. There remain aspects which are not within the framework of the advances in question, such as the rather rapid asymmetry development in the Clayton et al. (1957) experiments. A simple scale effect does not appear entirely adequate as an explanation, and no other explanation offers itself as a clear candidate. There still remain problems in the interpretation of selection experiments to provide challenges to theoreticians.

REFERENCES

BULMER, J.G. 1971. Am.Nat. 105: 201-211.
BULMER, M.G. 1974. Genet.Res. 23: 281-189.
CLAYTON, G.A., MORRIS, J.A. and ROBERTSON, A. 1957. J.Genet. 55: 131- 151.
FALCONER, D.S. 1973. Genet.Res. 22: 291-321.
FRANKHAM, R., JONES, L.P. and BARKER, J.S.F. 1968a. Genet.Res. 12: 237-248.
FRANKHAM, R., JONES, L.P. and BARKER, J.S.F. 1968b. Genet.Res. 12: 267-283.
HILL, W.G. 1976. Biometrics 32: 889-902.
HILL, W.G. 1977. Biometrics 33: 703-712.
HILL, W.G. 1982. Proc.Nat.Acad.Sci.USA. 79: 142-145.
JAMES, J.W. 1971. Genet.Res. 16: 241-250
JONES, L.P., FRANKHAM, R. and BARKER, J.S.F. 1968. Genet.Res. 12: 249-266
LYNCH, M. 1988. Genet.Res. 51: 137-148.
MONTEIRO, L.S. and FALCONER, D.S. 1966. Anim.Prod. 8: 179-192.
RAWLINGS, J.O. 1976. Biometrics 32: 875-887.
ROBERTSON, A. 1960. Proc.Roy.Soc.London B. 153: 234-249.
ROBERTSON, A. 1961. Genet.Res. 2: 189-194.
ROBERTSON, D.E. 1969. Ph.D. Thesis. University of New South Wales.
SHERIDAN, A.K. 1988. Anim.Breed.Abstr. 56: 877-889.
WILSON, S.P. 1974. Genetics 76: 823-836.
WRAY, N.R. 1989. Ph.D. Thesis. University of Edinburgh.
YOO, B.H. 1980. Theor.Appl.Genet. 57: 27-32.