

## APPENDIX

PREDICTING PERFORMANCE IN  $F_1$  AND  $F_2$  GENERATIONS

NEIL GILBERT

1. This appendix considers the performance of  $F_1$  families in the  $18 \times 18$  diallel cross and of their  $F_2$  progenies. These  $F_2$  families were grown in 1958 in a lattice square with two replicates and ten plants per plot. Parents were included in the experiment; thus they occurred twice in the same field, once with the  $F_1$ 's (six plants each) and once with the  $F_2$ 's (20 plants each). The parents  $\times$  experiments interaction did not signify, yet there was a considerable difference in average performance per plant. It is therefore not always permissible to lump these two sets of parents together.

2. *Constancy of  $F_1$  families between years.* In 1957 the  $F_1$  were grown in a glasshouse; in 1958 out of doors in unusually bad conditions. The analyses of variance are:

	Degrees of freedom	Mean Square for		
		Yield	Fruit number	Average fruit weight
Average main effects . . . . .	17	764	203.9	4.636
Main effects $\times$ years . . . . .	17	71	15.3	0.638
Average interactions . . . . .	135	77	11.8	0.225
Interactions $\times$ years . . . . .	135	68	11.5	0.156
Average parents . . . . .	17	161	46.5	1.022
Parents $\times$ years . . . . .	17	75	16.2	0.218
Error . . . . .		37	4.9	0.126

The main effects show a very satisfactory stability. Their interactions with years are of the same order of magnitude as their interactions with each other.

3. *Top-parent regression.* It is clear from table 3 that the mean values show general phenotypic dominance of top-parent. A multiple regression on top- and bottom-parents may, therefore, give better prediction than simple regression on mid-parents. It should be remembered that we are no longer dealing with general mean differences but with deviations from the general means. The results are contradictory. In every case except 1958  $F_1$  yield and fruit number, the regression on mid-parent (with one degree of freedom) accounts for all but an insignificant part of the multiple-regression sum of squares and a regression on top-parent alone gives relatively poor prediction. In the two cases mentioned, however, the situation is reversed and the top-parent on its own gives the best prediction although the mid-parent is still a good predictor (*i.e.* the M.S. for regression on mid-parent is not far short of that for regression on top-parent). Consequently it is best to use mid-parents (and, therefore, main effects) in predicting the relative merits of crosses. The values of these mid-parent regression coefficients will be discussed below.

4. *Main effects.* The main effect, or general combining ability, measures the average performance of a variety as a parent. Of course, what the breeder really wants to know is the contribution that a variety makes to the best of, say, its  $F_4$  descendants. From the algebraic identities given by Gilbert (1958), p. 483, it will be seen that the regression coefficient of main effect on parental yield equals

one-half that of the yield of each cross on its mid-parent, whatever the actual values of these regressions. Since a cross gets half its genotype from each parent the regression of main effect on parental value might be expected to equal one-half, apart from some environmental dilution. In the previous paper I stated that this regression "does not, on average, differ from half". The present data contradict this.

	Regressions of main effects on parental values		
	Yield	Fruit number	Average fruit weight
1957 F <sub>1</sub>	0.336 (±0.0297)	0.500 (±0.0243)	0.421 (±0.0263)
1958 F <sub>1</sub>	0.231 (±0.0370)	0.246 (±0.0237)	0.416 (±0.0294)
1958 F <sub>2</sub>	0.302 (±0.0226)	0.280 (±0.0106)	0.557 (±0.0403)

Now the above regression coefficients will be less than their "genetic" values owing to dilution by environmental variation. "Genetic" regressions were obtained by correcting for this (assuming that the "genetic" and "error" components of variance are additive), just to see what happened. It must be pointed out, however, that the error variance is shown below to change from variety to variety. The regressions still did not come up to expectation. When comparing main effects with parental values, therefore, it is necessary to abandon the "heterogeneity of potency" method which assumes a regression coefficient of one-half, and substitute a regression analysis. The main effects and parents are then compared in a relative, rather than an absolute, way. It will be seen below that the "heterogeneity of potency" idea is still of some value in another connection.

The situation can be summed up as follows; in the whole diallel cross, the mean of the hybrids usually exceeds the parental mean. In their deviations from their mean, the individual hybrids follow the corresponding mid-parents; but these deviations of the hybrids, although closely correlated with the deviations of the mid-parents from the parental mean, are not so large. The hybrids are distributed relatively compactly in the upper part of the parental range. Thus although the poorer parents show heterosis, the best parents do not.

In 1958 the F<sub>2</sub>, F<sub>1</sub> and parents were all grown in the same field. Appealing to the fact that, on average, the amount of heterozygosity in the F<sub>2</sub> is half that in the F<sub>1</sub>, it might be expected that F<sub>2</sub> main effect = ½ F<sub>1</sub> main effect + ½ (½ parental value). The multiple regression of F<sub>2</sub> main effect on F<sub>1</sub> main effect and parent gives :

$$\begin{array}{rcl}
 \text{Yield} & . & . & . & . & 0.665t + 0.666p/2 \\
 \text{Fruit number} & . & . & . & . & 0.312t + 0.287p/2 \\
 \text{Average fruit weight} & . & . & . & . & 0.523t + 0.663p/2
 \end{array}$$

Here *t* and *p* are themselves highly correlated, and reduction of their variances by subtracting the "error" component makes nonsense of these multiple regressions; in other words, the attempt to calculate a "genetical" multiple regression is a failure. The multiple regressions quoted show that, as before, the coefficients may differ from half. However, *t*+*p*/2 gives good relative prediction of the F<sub>2</sub> main effect.

None of these regressions show any sign of curvature. Altogether, the main effects behave satisfactorily, although not quite so well as might be hoped.

5. *Interactions.* The interaction, or specific combining ability for a particular cross, is the remainder when the general mean *m* and the two relevant main effects *t* have been subtracted from the F<sub>1</sub> yield. Since real interactions do occur, it is

interesting to see how constant they are from year to year and generation to generation. The correlations observed (each with 134 d.f.) are :

	Yield	Fruit number	Average fruit weight
1957 $F_1 \times 1958 F_1$ .	0.069	0.017	0.179
1957 $F_1 \times 1958 F_2$ .	0.111	0.106	0.218
1958 $F_1 \times 1958 F_2$ .	0.152	0.062	0.019

Subtraction of the relevant error components does not increase the correlations to the region of 0.6 which may be considered the minimum value necessary for prediction in plant breeding. Furthermore, in practice error is always with us and is unlikely to be reduced sufficiently for the "genotypic" component of interaction to be estimated at all precisely. Interactions apparently vary so much from year to year and generation to generation that they can be forgotten when deciding on the best parents.

Since, on average, the genotype of a cross is composed of half the genotypes of each parent, we can tentatively call the difference between parental yield  $p$  and  $m+2t$  ( $m$  = general parental mean,  $t$  = main effect for that parent) an interaction. From Gilbert (1958) it appears that these differences, labelled "heterogeneity of potency", are no more variable than the interactions. If this were not the case, we should have to ascribe some extraordinary property to the very condition of inbreeding, quite apart from the difference between the parental and  $F_1$  averages. These "heterogeneities of potency" are more accurately estimated than the ordinary interactions. They give the following correlations (each with 16 d.f.) :

	Yield	Fruit number	Average fruit weight
1957 $F_1 \times 1958 F_1$ .	-0.030	-0.184	0.391
1957 $F_1 \times 1958 F_2$ .	-0.198	-0.475	0.365
1958 $F_1 \times 1958 F_2$ .	0.454	0.613	0.258

Bearing in mind the differences in the experimental conditions in 1957 and 1958 and also that main effects (covering all crosses involving one parent) are intrinsically more interesting than interactions (involving one cross only) it appears that for purposes of prediction, interactions are best forgotten but that "heterogeneities of potency" cannot yet be so dismissed.

In connection with interactions the technique of constant-parent regression introduced by Hull (1946) might be mentioned. For each parent, the regression coefficient of the yields of the crosses on the yields of the respective second parents is calculated. Selves are usually included in this calculation. Since the yield of each cross is equal to  $m+t_1+t_2+t_{12}$  where  $m$  is the general  $F_1$  mean,  $t_1$  and  $t_2$  are main effects and  $t_{12}$  is the interaction, the constant-parent regression of this yield on  $p_2$  (parental yield) is equal to the regression of  $(t_2+t_{12})$  on  $p_2$  and so the sum of the regressions of  $t$  and  $t_{12}$  on  $p$ . I have already discussed the regression of the main effects  $t$  on the parents  $p$ . The  $k$  regression coefficients corresponding to the  $k$  constant-parents are sometimes statistically heterogeneous. This can be interpreted in genetic terms as a rather complicated effect of dominance and frequency

of additive genes (Griffing, 1950). For our purposes, however, such heterogeneity indicates significant constant-parent regressions of interactions on second-parent yields. This is not very surprising, for, on commonsense grounds two high-yielding parents can be expected to interact negatively rather than positively so that the interactions of crosses involving one high-yielding parent may be expected to be negatively correlated with the yield of the second parent. Since interactions are valueless for prediction anyway, such behaviour is of little interest. This does not of course discredit the constant-parent regression technique for more theoretical purposes.

6. *F<sub>2</sub> variances.* In the F<sub>2</sub>, each cross was represented by two plots of ten plants each, giving 18 d.f. for "error". Variances, when considered as statistics in their own right (rather than as indicators of the accuracy of means) are notoriously erratic, yet these figures show some interesting features. In such work it is usual to analyse log variance but in this case the use of the standard deviation is to be preferred since this is the statistic used in the estimation of upper percentiles. The results are substantially the same anyway, as may be expected. It is often argued that the differences between plants of identical genotype must be purely environmental, so that the "error" variance within different inbred lines should be constant. The tomato parents are all believed to be highly inbred commercial varieties, yet both parents and F<sub>2</sub> families show heterogeneity of variance ( $\chi^2(17 \text{ d.f.}) = 82.2$  and  $\chi^2(152 \text{ d.f.}) = 334.8$  by Bartlett's test). The average F<sub>2</sub> variance exceeds the average parental variance :

	Degrees of freedom	Yield Mean Square
F <sub>2</sub> variances . . .	2754	18.26
Parental variances . .	324	13.73

This variance ratio cannot be expected to follow Fisher's distribution exactly, since we have demonstrated heterogeneity, but it suffices to show that F<sub>2</sub>'s tend to be more variable than the parents. The F<sub>2</sub> standard deviations, when analysed in the usual way, themselves show "main effects" :

	Degrees of freedom	Mean Square
Main effects . . .	17	43.60
Interactions . . .	135	16.60

and these main effects are correlated with the corresponding parental standard deviations ( $r = 0.52$ , 16 d.f.). This situation can be partly—but by no means completely—explained away by the observed correlation between variance and mean yield. There is therefore some evidence that the parents transmit variability, as well as average performance, to their offspring. It is not certain whether this is purely a case of different genotypes having different error variance similar to that mentioned by Haldane (1957) quoting S. K. Roy, or whether there is also some residual heterozygosity in the "inbred" tomato varieties.