GENOTYPE $\times$ ENVIRONMENT INTERACTIONS

II. SOME GENETICAL CONSIDERATIONS

KENNETH MATHER

Department of Genetics, University of Birmingham

Received 23.x.74

SUMMARY

An algebraic formulation, alternative to that of Mather and Jones (1958) and hierarchical rather than factorial in nature, is presented for describing the differences among the phenotypes produced by a number of genotypes each grown in each of a number of environments. This formulation does not include terms representing statistical interactions between genotypes and environments: it depends instead on comparisons between the different genotypes in their variation over the relevant ranges of environments.

The two-line case is considered and the condition established for linearity of the regression of genotype $\times$ environment interactions (g in Mather and Jones' formulation) on overall effect of the environment (e in Mather and Jones' formulation). This condition is the same as that derived in a different way by Mather and Caligari. It is shown that genes which respond to the environmental differences but which do not differ between the genotypes under observation will nevertheless affect the relation of genotype $\times$ environment interaction to the overall effect of the environment. In particular, different values may be obtained for the regression of g on e solely because of the effects of genes in which the genotypes under comparison do not differ.

The multi-line case is developed as a simple extension of the two-line situation. Again the condition for linearity of the regressions is obtained and shown to be the same as that derived earlier and differently by Mather and Caligari. The value of genetically non-identical, but closely related, material for the biological measurement of the environment is discussed, and is shown to depend on the variation in the responses of the various genes to the environmental changes. Where these responses are similar any reasonably related material will give an adequate representation; but where the various genes may respond differently, the choice of material for the biological measurement of the environment is much more critical.

Multiplicative action of genotype and environment is considered. Not only must the plot of g for each genotype against the general e give a set of straight lines passing through the origin, but the plot of the phenotype from each genotype against the average of all phenotypes, environment by environment, must also give a set of straight lines passing through the origin. Thus multiplicative action should not be difficult to detect by routine biometric analysis.

The environment is made up of factors which may vary independently of one another and to whose changes the different genes affecting a character will not respond equally. The case is considered of two environmental factors and two gene-pairs, one reacting to each of the factors. The condition is derived for a straight regression line of g on e . It is shown that if one pair of genotypes, say AB and ab, give a straight regression line, the other genotypes, Ab and aB, will give a box like relation of g to e . Thus recombination converts a straight regression line into a regression box (fig. 2). This basic result still obtains where sets of genes rather than single genes react to change in an environmental factor and also where genes can be pleiotropic in their responses in that they respond to changes in both environmental factors.

Finally, the consequences are considered of incomplete sets of genotypes being observed in incomplete sets of environments. Use of an incomplete set of environments does not destroy any straight regression line that may exist

for g on e . With an incomplete set of genotypes, the straight regression may be preserved or it may be lost according to which genotype is missing; but where it is lost a near straight regression line is obtained for another genotype (fig. 4). In short a linear relation of g to e will seldom be lost or even obscured by failure to raise all possible genotypes in all possible environments.

1. INTRODUCTION

IN their discussion of genotype $\times$ environment interactions Mather and Jones (1958) derived an algebraic formulation which has been used extensively in later analyses. This formulation is set out in table 1a for the simplest case of two true-breeding genotypes, differing in respect of one gene raised in two environments. Three parameters are required to describe the differences among the four phenotypes and these are chosen as d , the genetic parameter measuring the difference between the two genotypes when averaged over both environments; e , the environment parameter

TABLE 1

The old formulation of Mather and Jones (a) and the present alternative formulation (b) of the three parameters required completely to describe the differences among the four phenotypes given by two genotypes in two environments

		Environments	
		1	2
(a)	Genotypes		
	AA	$d - e - g$	$d + e + g$
	aa	$-d - e + g$	$-d + e - g$
		Environments	
		1	2
(b)	Genotypes		
	AA	$d - e_A$	$d + e_A$
	aa	$-d - e_a$	$-d + e_a$

measuring the difference between the two environments averaged over both genotypes; and g , the parameter measuring the interaction of the genetical and environmental effects found as the statistical interaction between d and e . The sum of squares of differences among the four phenotypes is then divided into three items, the genetical springing from d , the environmental springing from e , and the interactive springing from g . This formulation follows the pattern most commonly used in the analysis of variance of a simple two-way factorial situation. Being factorial, it is of course symmetrical in its representation of the genetical and environmental elements of the phenotypic differences.

Other formulations of the three parameters necessary for a full description of the data are possible and one of them is set out in table 1b. The three parameters are now d , which as before measures the difference between the two genotypes averaged over environments; e_A which measures the difference in genotype AA between the two environments; and e_a which similarly measures the difference in genotype aa between the two environments. Like d , e and g , the three parameters of this new set are orthogonal

to one another and lead to an analysis of variance into three parts depending respectively on the overall or average genetical difference, the difference produced in AA by the change of environment and the difference similarly produced in aa by the change of environment. This formulation is, of course, not factorial: on the contrary it is hierarchial. The genetical and environmental effects are no longer treated symmetrically, and interaction in the statistical sense no longer appears directly in the formulation. Instead, attention is focused on the effect of the environmental difference by the way in which each individual genotype manifests itself. The formulation might thus be held to be more appropriate to a consideration of the role of the genes in their interplay with the environment.

It should be observed right away that d , e_A and e_a among them contain exactly the same information as d , e and g . In fact, d is the same in both cases and $e = \frac{1}{2}(e_A + e_a)$ with $g = \frac{1}{2}(e_A - e_a)$ or, the other way round, $e_A = e + g$ and $e_a = e - g$. So any result obtained from one formulation can equally be obtained from the other or re-written in terms of it. Nevertheless, certain results may be more readily established, or their implications made more immediately obvious in terms of one formulation rather than the other.

2. TWO-LINE CASE

The simple case of two genotypes differing in one gene and raised in two environments is readily extended to cover any number of genes and any number of environments. Adapting the notation of Perkins and Jinks (1968a), the phenotypes given by the two genotypes raised over s environments can be represented as in table 2. The overall genetical

TABLE 2
The formulation for two genotypes over s environments

Genotypes	Environments				Mean
	1	2	...	s	
1	$d_1 + e_{11}$	$d_1 + e_{12}$	...	$d_1 + e_{1s}$	d_1
2	$d_2 + e_{21}$	$d_2 + e_{22}$	...	$d_2 + e_{2s}$	d_2
Mean	e_1	e_2	...	e_s	0

$$\text{where } d_1 = -d_2, \sum_1^s (e_{1j}) = \sum_1^s (e_{2j}) = 0$$

component of the variation among the phenotypes depends on d_1 and d_2 ($= -d_1$ in the two-line case) while e_{1j} and e_{2j} represent the responses of genotypes 1 and 2 respectively to the j th environment. The "effect" of the first environment, found by averaging the phenotypes given by the two genotypes in this environment, is $e_1 = \frac{1}{2}(e_{11} + e_{21})$ and that of the j th environment is $e_j = \frac{1}{2}(e_{1j} + e_{2j})$. The element of genotype $\times$ environment interaction of genotype 1 in environment j will be given by the difference between the phenotype observed, viz. $d_1 + e_{1j}$, and that expected from the margins of the table, viz. $d_1 + e_j$. This gives $g_{1j} = d_1 + e_{1j} - d_1 - e_j = \frac{1}{2}(e_{1j} - e_{2j})$. Similarly the interaction component of genotype 2 in environment j is

$g_{aj} = \frac{1}{2}(e_{2j} - e_{1j})$. The interaction associated with environment j thus depends on $g_j = \frac{1}{2}(e_{1j} - e_{2j})$. Then the regression of g on e , which is widely used to measure the relationship between the genotype $\times$ environment interaction and the overall effect of the environment is

$$b_{g/e} = \frac{W_{g/e}}{V_e} = \frac{W_{(e_{1j}-e_{2j})(e_{1j}+e_{2j})}}{V_{(e_{1j}+e_{2j})}},$$

which is a measure of the rate of change of $e_{1j} - e_{2j}$ on $e_{1j} + e_{2j}$, though of course biased by the inflationary effects of error variation on $V_{(e_{1j}+e_{2j})}$.

Now the values of e_{1j} , e_{2j} , etc., will be affected by all the genes carried by the two genotypes whose actions are affected by the differences among the environments. Considering locus A, its contribution to e_{1j} will be e_{Aj} and to e_{2j} will be e_{aj} , where A is the allele in genotype 1 and a that in genotype 2. Its contribution to e_j will thus be $\frac{1}{2}(e_{Aj} + e_{aj})$ and to g will be $\frac{1}{2}(e_{Aj} - e_{aj})$, or summing over all genes involved $e = \frac{1}{2}S(e_A + e_a)$ and $g = \frac{1}{2}S(e_A - e_a)$. Any gene whose expression is affected by the environmental differences will thus contribute to e , but it will contribute to g only if $e_A - e_a \neq 0$, i.e. only if the alleles in the two genotypes differ in their responses to the environments: if $e_A - e_a = 0$, g will not be affected. This may come about in two ways. The alleles may differ in their overall effects on the phenotype, i.e. $d \neq 0$, but respond in the same way to change in the environment, i.e. $e_A = e_a$, etc. Or the alleles may have the same effect on the phenotype as one another in all environments, i.e. $d = 0$ as well as $e_A = e_a$. In this case the genotypes are carrying identical alleles at locus A, at any rate so far as these observations are concerned. The genes affecting the phenotype which do not differ between the genotypes will contribute to the value of e_j but not to the value of g_j , which can be affected only by genes in which the two genotypes differ. This confirms our intuitive expectation that genes which do *not* differ between the genotypes, can affect the biological assessment (e) of the environment; but they obviously cannot produce any genotype $\times$ environment interaction (g).

Genes in which the genotypes do not differ will therefore affect the regression of g on e , as well as genes in which the genotypes do differ. In particular, two pairs of genotypes, X_1 and X_2 being one pair and Y_1 and Y_2 the other, need not give the same values for the regression of g on e even though Y_1 differed from Y_2 in exactly the same genes as X_1 differed from X_2 , if the background genes common to Y_1 and Y_2 were not the same as those common to X_1 and X_2 . There is evidence from *Drosophila melanogaster* of just such differences in the regression of g on e arising from changes in the background genes (Caligari and Mather, unpublished).

Turning to the form of the regression of g on e , it will be simply linear if, in respect of any four environments 1 to 4

$$\frac{g_1 - g_2}{e_1 - e_2} = \frac{g_3 - g_4}{e_3 - e_4}.$$

Re-writing this in terms of the responses of the individual genes, e_1 , etc., gives

$$\frac{e_{11} - e_{21} - e_{12} + e_{22}}{e_{11} + e_{21} - e_{12} - e_{22}} = \frac{e_{13} - e_{23} - e_{14} + e_{24}}{e_{13} + e_{23} - e_{14} - e_{24}},$$

which reduces to

$$\frac{e_{11} - e_{12}}{e_{21} - e_{22}} = \frac{e_{13} - e_{14}}{e_{23} - e_{24}} = k.$$

The numerators refer to changes in expression of the genes in genotype 1 and the denominators to changes in expression of their alleles in genotype 2. Thus the regression of g on e will be rectilinear if, and only if, the overall change in response of the genes in genotype 1, brought about by change in the environment, bears a constant ratio k to the overall change in response of their alleles in genotype 2. This condition for linearity, now set out in terms of specific gene actions, is the same as that derived in a different way by Mather and Caligari (1974); and as these authors observed, it shows that such a linear regression of gene neither implies nor requires linear regression of the phenotypic expression on any physical measurement of the factors by which the environment is changing, such as temperature, humidity, etc. It requires only that the genotypes respond in basically the same way, however complicated this may be when related to the external or physical measurement.

In so far as the two genotypes are responding unequally to changes in the environment, the variances of their phenotypes over environments, viz. $V_{e_{1j}}$ and $V_{e_{2j}}$, will be unequal, apart from the special case where e_{1j} is changing by increments equal in size but opposite in sign to those of e_{2j} . Differences in the responses of gene expression to environment, and the genotype $\times$ environment interactions to which they lead, can thus be detected by inequality of the variances of the genotypes over any given range of environments. Furthermore, if the relationship between their changes is of the simple kind that leads to a rectilinear regression of g on e , the ratio of the two variances, $V_{e_{1j}}$ and $V_{e_{2j}}$ (corrected by deduction of a component representing random or error variation not accounted for by reference to the environment as measured in the experiment) will be k^2 . Thus comparison of the genotype variances can provide the same type of information as the regression analysis, at least in the case of a simple relationship. Whether, however, it could deal as well with more complex relations requires further investigation.

3. MULTI-LINE CASE

Turning to the case of t genotypes raised in s environments, as described by Perkins and Jinks (1968a), the formulations may be set out as in table 3. Since $S(d_i) = 0$, $e_1 = S(e_{i1})$, which with $S(e_{1j}) = 0$, etc., gives $S(e_1) = 0$. The interactive element in the phenotype of genotype 1 in environment 1 will thus be

$$\begin{aligned} g_{11} &= d_1 + e_{11} - d_1 - e_1 = e_{11} - \frac{1}{t} S(e_{i1}) \\ &= \frac{1}{t} [(e_{11} - e_{11}) + (e_{11} - e_{21}) \dots + (e_{11} - e_{t1})] \\ &= \frac{1}{t} \sum_{i=2}^t (e_{11} - e_{i1}). \end{aligned}$$

Then following Perkins and Jinks, we can denote by β_1 the regression of g_{11} , etc., the interactive element of genotype 1 on e_1 , etc., the overall environmental effects. β_1 is thus the regression of

$$\frac{1}{t} S(e_{11} - e_{i1}) \text{ on } \frac{1}{t} S(e_{i1}).$$

So in parallel with two line case, β_1 is the regression of the average of the differences between the responses of genotype 1 and the other $t-1$ genotypes taken in turn, on the summed responses of all t lines to that environment. Again the values of e_1 , etc., will reflect the action of genes in which

TABLE 3
The formulation for t genotypes over s environments

Genotypes	Environments				Mean
	1	2	...	s	
1	$d_1 + e_{11}$	$d_1 + e_{12}$	...	$d_1 + e_{1s}$	d_1
2	$d_2 + e_{21}$	$d_2 + e_{22}$	...	$d_2 + e_{2s}$	d_2
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
t	$d_t + e_{t1}$	$d_t + e_{t2}$	...	$d_t + e_{ts}$	d_t
Mean	e_1	e_2	...	e_s	0

$$\text{where } \sum_1^t (d_i) = \sum_1^s (e_j) = 0$$

the t genotypes do not differ as well as of the genes in which they do, while g_{11} , etc., will reflect only the action of genes by which the genotypes do differ. Thus β_1 will depend for its value on the constant or background genes, which are the same in all t lines, as well as the varying genes which are not.

The condition for rectilinearity of the regression is readily found to be

$$\frac{\frac{e_{11} - e_{12}}{2}}{\frac{\sum_1^t (e_{i1}) - \sum_2^t (e_{i2})}{2}} = \frac{\frac{e_{13} - e_{14}}{2}}{\frac{\sum_1^t (e_{i3}) - \sum_2^t (e_{i4})}{2}} = k,$$

which is clearly a simple extension of the condition already found for the two-line case. The regression of g on e will thus be rectilinear if the overall response of the genes in genotype 1 brought about by change of environment bears a constant ratio k to the sum of the overall responses of the genes in the remaining genotypes.

A further point may be made about the regression of g on e . There has been discussion in the literature, starting with Freeman and Perkins (1971), about the desirability of involving the same observations in the estimation of g and e , and about the possible need for including special material for the estimation of e , such material might be a duplicate of that from which g is found, or alternatively might be a special sample of the genotypes under investigation or of related genotypes. The putative statistical difficulty from which the discussion originated has been found to be unreal, or at least avoidable (Mather and Caligari, 1974), but a different

aspect of the situation is worthy of note. The value of e_i depends on the sum of the responses to the environment of all the relevant genes of all the genotypes, including both the constant and the variable genes. Now if all, or a great preponderance, of the genes, whether at different loci in the case of constant genes or alleles in the case of variable genes, react in the same direction and broadly in the same order of magnitude to the changes of the environment, e will change much more and much more nearly uniformly with the environment than will g , which will reflect only the differences in response of the alleles at the variable loci. Under these circumstances it will not greatly affect the regression of g on e whether the same genotypes or other reasonably related lines are used for estimating e or for estimating the various g 's, though no doubt identical genotypes will give a somewhat better fitting regression. This has indeed been found experimentally to be the case by Fripp and Caten (1971) and Fripp (1972) who investigated the response in rate of growth of the fungus *Schizophyllum commune* to environments differing in temperature and composition of the medium. They found that although there was a tendency for the regression to become somewhat less satisfactory as the genotypes used for estimating e became more distantly related to those used to provide the g 's, the regression obtained using any of the various assessments of e were in fact remarkably similar.

The situation is quite different when we turn to the effect of temperature and cultural condition on the number of sternopleural chaetae in *Drosophila melanogaster*. In most genotypes the number of chaetae rises with temperature, but in a few it falls, though by a relatively small amount. It is thus clear that there must be genes, and allelic genes, which respond in opposite directions to such change in the environment (see Mather and Caligari, 1974). The changes in e will thus reflect the resultant of the genes, both constant and variable, which are acting, and change in e can and does change in sign as well as magnitude according to the set of genotypes used for its assessment. Furthermore, when genotypes responding in opposite direction are used, g can be greater than e and can thus give a regression line of slope above 1 or below -1 according to which genotype is its source. The broad general validity of the regressions obtained by Fripp and Caten no longer apply in such a case: the choice of genotypes for assessing the environment is of central importance because the genes affecting the character no longer preponderantly affect it in a broadly similar way.

4. MULTIPLICATIVE ACTION

The relation between genotype and environment may be said to be multiplicative where the increment added by change of environment is proportional to the expression of the character in the genotype. Thus a fertiliser treatment might be found to raise the yield of each of a series of plant varieties by a proportionate amount—by 10, 50 or 100 per cent or whatever it may be. The differences among the varieties would thus be increased in absolute terms by the change of environment, but they would still be the same in proportionate terms. In such a case the phenotype can be expressed as the product xy , where x is the genetic component and y the environmental. The case of t varieties in s environments is set out in

table 4. Clearly if we transform the data by taking logarithms, $x_i y_j$ being replaced by $\log x_i + \log y_j$, the two components become simply additive

TABLE 4
Multiplicative action of genotype and environment

Genotypes	Environments				Mean
	1	2	...	s	
1	$x_1 y_1$	$x_1 y_2$	...	$x_1 y_s$	$x_1 \bar{y}$
2	$x_2 y_1$	$x_2 y_2$	...	$x_2 y_s$	$x_2 \bar{y}$
.	.	.		.	.
.	.	.		.	.
t	$x_t y_1$	$x_t y_2$	...	$x_t y_s$	$x_t \bar{y}$
Mean	$\bar{x} y_1$	$\bar{x} y_2$	...	$\bar{x} y_s$	$\bar{x} \bar{y}$

and there is no genotype $\times$ environment interaction to take into account or analyse further. If, on the other hand, the untransformed data are analysed as they stand, genotype $\times$ environment interactions will appear. They nevertheless follow a characteristic pattern by which multiplicative action may be readily recognised.

The genotype means and environment means are shown in the margins of table 4. If we plot the phenotypes of any genotype, environment by environment, against the mean of all genotypes in the relevant environment, we obtain a straight regression line: the change in phenotype of genotype i in moving from environment 1 to environment 2 will be $x_i(y_2 - y_1)$ while the mean of all genotypes changes by $\bar{x}(y_2 - y_1)$ which gives

$$\frac{x_i(y_2 - y_1)}{\bar{x}(y_2 - y_1)} = \frac{x_i}{\bar{x}}$$

as the slope of the regression line, no matter what values y_1 and y_2 may have. Furthermore, this regression line must pass through the origin since the expression of the character by genotype i will be 0 when y is 0 and all the genotypes, and hence also their mean, will thus equally have 0 expression. Thus a characteristic pencil of straight lines would be obtained, the line given by genotype i having a slope of $\frac{x_i}{\bar{x}}$, and all the lines passing through the origin (fig. 1).

If we proceed to a formal biometrical-genetical analysis, using Mather and Jones' parameters, we set the expression of the i th genotype in the j th environment at

$$x_i y_j = m + d_i + e_j + g_{ij},$$

where m is the mean of the whole set of observations.

Now d_i is the deviation from this grand mean of the i th genotype averaged over all environments, and e_j is the deviation from the grand mean of the j th environment averaged over all genotypes. Thus from table 4

FIG. 1.—Multiplicative action of genotype and environment. *Above*: Phenotypes of individual genotypes plotted against the mean of all genotypes, environment by environment. *Below*: Genotype $\times$ environment interaction (g) plotted against overall effect of the environment (e). Three genotypes are shown in four environments, the genotypes having values of $x = 5, 2, 1$ and the environments values of $y = 1, 3, 4, 6$ (see in text).

$$\begin{aligned}
 m &= \bar{x}\bar{y} \\
 d_i &= (x_i - \bar{x})\bar{y} \\
 e_j &= \bar{x}(y_j - \bar{y})
 \end{aligned}$$

and

$$g_{ij} = x_i y_j - \bar{x}\bar{y} - (x_i - \bar{x})\bar{y} - \bar{x}(y_j - \bar{y}) = (x_i - \bar{x})(y_j - \bar{y}).$$

Then regressing g on e we find the slope of the line for genotype i as

$$\begin{aligned}
 \frac{g_{i1} - g_{i2}}{e_1 - e_2} &= \frac{(x_i - \bar{x})(y_1 - \bar{y}) - (x_i - \bar{x})(y_2 - \bar{y})}{\bar{x}(y_1 - \bar{y}) - \bar{x}(y_2 - \bar{y})} \\
 &= \frac{(x_i - \bar{x})(y_1 - y_2)}{\bar{x}(y_1 - y_2)} = \frac{x_i - \bar{x}}{\bar{x}}.
 \end{aligned}$$

The slope is thus independent of y and the regression line must be straight, and all lines will pass through the origin since by definition this marks the mean of e and the mean of g (fig. 1, below). This pattern of lines, all of them necessarily straight, is very distinctive. Together with the pencil of straight lines passing through the origin when the phenotype of each genotype is plotted against the mean of all genotypes, it should provide a means of recognising multiplicative action with little difficulty when the customary type of genotype $\times$ environment interaction analysis is applied.

The algebraic relation between the genetical and environmental components is symmetrical in multiplicative action. We may note therefore that similar sets of necessarily straight lines would be obtained by regression of g on d instead of e , and also by regressing genotype by genotype, the expression of each environment on the mean of all environments.

5. FACTORS OF THE ENVIRONMENT

The environment is made up of many factors whose differences and interplay result in the complex variation of the environment itself. These variable factors may be of many different kinds and we cannot expect the genes affecting the manifestation of a character all to respond equally to all their changes: some genes will react chiefly to changes in certain of the factors and other genes will react more to changes in other factors, though of course the expression of any gene may be affected in its own particular way by changes in more than one factor.

Let us consider the simplest case of two environmental factors, α and β , each at two levels, 1 and 2, and combining in all possible ways to give four environments, $\alpha_1\beta_1$, $\alpha_2\beta_1$, $\alpha_1\beta_2$ and $\alpha_2\beta_2$. Let us further assume that there are two gene pairs, A-a and B-b, which react specifically to change in α and β respectively, there being four genotypes AB, Ab aB and ab. The parameters measuring the reaction of A and a to the change in α are e_A and e_a respectively, and those measuring the reaction of B and b to the change in β are similarly e_B and e_b . The phenotypes expected from the four genotypes in the four environments are set out in table 5, together with the overall values of e and the individual g 's obtained when all genotypes are raised in all environments. For the sake of both brevity and clarity we will write e_A as A , e_a as a , e_B as B and e_b as b in what follows.

TABLE 5

Two gene pairs, A-a and B-b, reacting specifically to two environmental factors α and β . The phenotypes are shown in the upper part and the g 's in the lower part of the table

Phenotypes	Environments				Mean	Variance
	$\alpha_1\beta_1$	$\alpha_2\beta_1$	$\alpha_1\beta_2$	$\alpha_2\beta_2$		
AB	$d_a + d_b$ $-A - B$	$d_a + d_b$ $+A - B$	$d_a + d_b$ $-A + B$	$d_a + d_b$ $+A + B$	$d_a + d_b$	$A^2 + B^2$
Ab	$d_a - d_b$ $-A - b$	$d_a - d_b$ $+A - b$	$d_a - d_b$ $-A + b$	$d_a - d_b$ $+A + b$	$d_a - d_b$	$A^2 + b^2$
aB	$-d_a + d_b$ $-a - B$	$-d_a + d_b$ $+a - B$	$-d_a + d_b$ $-a + B$	$-d_a + d_b$ $+a + B$	$-d_a + d_b$	$a^2 + B^2$
ab	$-d_a - d_b$ $-a - b$	$-d_a - d_b$ $+a - b$	$-d_a - d_b$ $-a + b$	$-d_a - d_b$ $+a + b$	$-d_a - d_b$	$a^2 + b^2$
Mean (= e)	$-\frac{1}{2}(A+a)$ $-\frac{1}{2}(B+b)$	$\frac{1}{2}(A+a)$ $-\frac{1}{2}(B+b)$	$-\frac{1}{2}(A+a)$ $+\frac{1}{2}(B+b)$	$\frac{1}{2}(A+a)$ $+\frac{1}{2}(B+b)$	0	
	$V_e = \frac{1}{4}[(A+a)^2 + (B+b)^2]$				V_g	W
g 's	$-\frac{1}{2}(A-a)$ $-\frac{1}{2}(B-b)$	$\frac{1}{2}(A-a)$ $-\frac{1}{2}(B-b)$	$-\frac{1}{2}(A-a)$ $+\frac{1}{2}(B-b)$	$\frac{1}{2}(A-a)$ $+\frac{1}{2}(B-b)$	$\frac{1}{4}[(A-a)^2 + (B-b)^2]$	$\frac{1}{4}(A^2 - a^2 + B^2 - b^2)$
	$-\frac{1}{2}(A-a)$ $+\frac{1}{2}(B-b)$	$\frac{1}{2}(A-a)$ $+\frac{1}{2}(B-b)$	$-\frac{1}{2}(A-a)$ $-\frac{1}{2}(B-b)$	$\frac{1}{2}(A-a)$ $-\frac{1}{2}(B-b)$	$\frac{1}{4}[(A-a)^2 + (B-b)^2]$	$\frac{1}{4}(A^2 - a^2 - B^2 + b^2)$
	$\frac{1}{2}(A-a)$ $-\frac{1}{2}(B-b)$	$-\frac{1}{2}(A-a)$ $-\frac{1}{2}(B-b)$	$\frac{1}{2}(A-a)$ $+\frac{1}{2}(B-b)$	$-\frac{1}{2}(A-a)$ $+\frac{1}{2}(B-b)$	$\frac{1}{4}[(A-a)^2 + (B-b)^2]$	$\frac{1}{4}(-A^2 + a^2 + B^2 - b^2)$
	$\frac{1}{2}(A-a)$ $+\frac{1}{2}(B-b)$	$-\frac{1}{2}(A-a)$ $+\frac{1}{2}(B-b)$	$-\frac{1}{2}(A-a)$ $-\frac{1}{2}(B-b)$	$-\frac{1}{2}(A-a)$ $-\frac{1}{2}(B-b)$	$\frac{1}{4}[(A-a)^2 + (B-b)^2]$	$\frac{1}{4}(-A^2 + a^2 - B^2 + b^2)$

Note.— $A = e_A$, $a = e_a$, $B = e_B$, $b = e_b$ (see in text).

The variance of e (found as the mean of all four genotypes in each environment) is

$$V_e = \left[-\frac{1}{2}(A+a) - \frac{1}{2}(B+b) \right]^2 + \left[\frac{1}{2}(A+a) - \frac{1}{2}(B+b) \right]^2 \dots$$

$$= \frac{1}{4}[(A+a)^2 + (B+b)^2].$$

This value can be obtained for V_e not only from all the four genotypes raised together but also when only the two genotypes AB and ab, or alternatively the two genotypes Ab and aB are raised in each environment. The phenotypic variances of the four genotypes over environments are shown on the right side of the upper part of table 5 and it will be seen that they differ from one another. The variance of the g 's shown on the right side of the lower part of the table, is the same for all four genotypes at

$$V_g = \frac{1}{4}[(A-a)^2 + (B-b)^2].$$

At first sight this is a surprising result, which stems however from the fact that, because of the relation between genotype and environments, all four genotypes generate the same four values of g but of course do so in different

TABLE 6
 V_e , V_g , W and remainder variance with genotypes and environments missing

Genotype	V_g	W	Remainder variance
<i>Four genotypes in four environments</i>			
AB, ab	$\frac{1}{4}[G_A^2 + G_B^2]$	$\pm \frac{1}{4}(E_A G_A + E_B G_B)$	$\frac{1}{4}(Ab - aB)^2/V_e$
Ab, aB	$\frac{1}{4}[G_A^2 + G_B^2]$	$\pm \frac{1}{4}[E_A G_A - E_B G_B]$	$\frac{1}{4}(AB - ab)^2/V_e$
<i>One environment missing</i>			
Symmetrical			
AB, ab	$\frac{1}{8}[G_A^2 + G_B^2 + (G_A - G_B)^2]$	$\pm \frac{1}{8}[E_A G_A + E_B G_B + (E_A - E_B)(G_A - G_B)]$	$\frac{4}{8}(Ab - aB)^2/V_e$
Ab, aB	$\frac{1}{8}[G_A^2 + G_B^2 + (G_A + G_B)^2]$	$\pm \frac{1}{8}[E_A G_A - E_B G_B + (E_A - E_B)(G_A + G_B)]$	$\frac{4}{8}(AB - ab)^2/V_e$
Asymmetrical			
AB, ab	$\frac{1}{8}[G_A^2 + G_B^2 + (G_A + G_B)^2]$	$\pm \frac{1}{8}[E_A G_A + E_B G_B + (E_A + E_B)(G_A + G_B)]$	$\frac{4}{8}(Ab - aB)^2/V_e$
Ab, aB	$\frac{1}{8}[G_A^2 + G_B^2 + (G_A - G_B)^2]$	$\pm \frac{1}{8}[E_A G_A - E_B G_B + (E_A + E_B)(G_A - G_B)]$	$\frac{4}{8}(AB - ab)^2/V_e$
<i>One genotype missing</i>			
Symmetrical			
AB	$\frac{1}{6}[G_A^2 + G_B^2]$	$\frac{1}{6}[E_{2A}G_A + E_{2B}G_B]$	$\frac{1}{6}(Ab - aB)^2/V_e$
Ab	$\frac{1}{6}[G_A^2 + 4G_B^2]$	$\frac{1}{6}[E_{2A}G_A - 2E_{2B}G_B]$	$\frac{1}{6}(2AB - Ab - ab)^2/V_e$
aB	$\frac{1}{6}[4G_A^2 + G_B^2]$	$\frac{1}{6}[-2E_{2A}G_A + E_{2B}G_B]$	$\frac{1}{6}(2AB - aB - ab)^2/V_e$
Asymmetrical			
AB	$\frac{1}{6}[G_A^2 + 4G_B^2]$	$\frac{1}{6}[E_{2A}G_A + 2E_{2B}G_B]$	$\frac{1}{6}(2A_b - AB - aB)^2/V_e$
Ab	$\frac{1}{6}[G_A^2 + G_B^2]$	$\frac{1}{6}[E_{2A}G_A - E_{2B}G_B]$	$\frac{1}{6}(AB - ab)^2/V_e$
ab	$\frac{1}{6}[4G_A^2 + G_B^2]$	$\frac{1}{6}[-2E_{2A}G_A - E_{2B}G_B]$	$\frac{1}{6}(2Ab - aB - ab)^2/V_e$

One genotype and one environment missing

	Sym/Sym	
AB	$\frac{4}{81}[C_A^2 + C_B^2 + (G_A - G_B)^2]$	$\frac{1}{243}(Ab - aB)^2/V_e$
Ab	$\frac{4}{81}[C_A^2 + 4C_B^2 + (G_A + 2G_B)^2]$	$\frac{1}{243}(2AB - Ab - ab)^2/V_e$
aB	$\frac{4}{81}[4C_A^2 + C_B^2 + (2G_A + G_B)^2]$	$\frac{1}{243}(2AB - aB - ab)^2/V_e$
	Sym/Asym	
AB	$\frac{4}{81}[C_A^2 + C_B^2 + (G_A + G_B)^2]$	$\frac{1}{243}(Ab - aB)^2/V_e$
Ab	$\frac{4}{81}[C_A^2 + 4C_B^2 + (G_A - 2G_B)^2]$	$\frac{1}{243}(2AB - Ab - ab)^2/V_e$
aB	$\frac{4}{81}[4C_A^2 + C_B^2 + (2G_A - G_B)^2]$	$\frac{1}{243}(2AB - aB - ab)^2/V_e$
	Asym/Sym	
AB	$\frac{4}{81}[C_A^2 + 4C_B^2 + (G_A - 2G_B)^2]$	$\frac{1}{243}(2Ab - AB - aB)^2/V_e$
Ab	$\frac{4}{81}[C_A^2 + C_B^2 + (G_A + G_B)^2]$	$\frac{1}{243}(AB - ab)^2/V_e$
ab	$\frac{4}{81}[4C_A^2 + C_B^2 + (2G_A - G_B)^2]$	$\frac{1}{243}(2Ab - aB - ab)^2/V_e$
	Asym/Asym	
AB	$\frac{4}{81}[C_A^2 + 4C_B^2 + (G_A + 2G_B)^2]$	$\frac{1}{243}(2Ab - AB - aB)^2/V_e$
Ab	$\frac{4}{81}[C_A^2 + C_B^2 + (G_A - G_B)^2]$	$\frac{1}{243}(AB - ab)^2/V_e$
ab	$\frac{4}{81}[4C_A^2 + C_B^2 + (2G_A + G_B)^2]$	$\frac{1}{243}(2Ab - aB - ab)^2/V_e$

The expressions for V_g , W and remainder variance are set out in the columns as shown by the headings at the top of the table; but note that the expressions for V_e are inserted separately in each section of the table.

Note.— $E_A = A + a = e_A + e_a$; $E_{2A} = 2A + a = 2e_A + e_a$

$E_B = B + b = e_B + e_b$; $E_{2B} = 2B + b = 2e_B + e_b$; $E_{2b} = B + 2b = e_B + 2e_b$

$G_A = A - a = e_A - e_a$; $G_B = B - b = e_B - e_b$.

It should be observed that for convenience E and G are here used to represent linear quantities. This is in contrast to the more general usage of biometrical genetics where E and G are used to denote components of variation, which are quadratic quantities.

environments. The covariance W of g with e taken over the four environments differs from one genotype to another, but it will be noticed that W for genotype ab is equal in magnitude but opposite in sign to that for AB , and W for aB is similarly related to that for Ab . The regression of g on e for genotype AB is thus

$$b_{AB} = (A^2 - a^2 + B^2 - b^2) / [(A+a)^2 + (B+b)^2]$$

and it accounts for $(A^2 - a^2 + B^2 - b^2)^2 / 4[(A+a)^2 + (B+b)^2]$ of the variance of g , leaving a remainder mean square $V_g - W^2/V_e = \frac{1}{4}(Ab - aB)^2/V_e$ (see table 6). Similarly for genotype Ab ,

$$b_{Ab} = (A^2 - a^2 - B^2 + b^2) / [(A+a)_2 + (B+b)^2]$$

accounting for $(A^2 + a^2 - B^2 + b^2)^2 / 4[(A+a)^2 + (B+b)^2]$ of V_g and leaving a remainder mean square of $\frac{1}{4}(AB - ab)^2/V_e$. Thus if

$$\frac{A}{a} = \frac{B}{b} = k, \quad Ab - aB = 0$$

and genotypes AB and ab will give straight regression lines, leaving no residual variation round the line. Similarly genotypes Ab and aB will give straight lines with no residual variations when

$$\frac{A}{a} = \frac{b}{B} \quad \text{giving} \quad AB - ab = 0.$$

The two pairs of genotypes cannot both give straight regression lines with no residual variation unless

$$\frac{A}{a} = \frac{a}{A} = \frac{B}{b} = \frac{b}{B} = 1$$

when $A = a$ and $B = b$, in which case there is no genotype $\times$ environment interaction.

The perfectly fitting regression given by genotypes AB and ab when $\frac{A}{a} = \frac{B}{b}$ is illustrated in fig. 2 (top) for the case where $A = 4$, $a = B = 2$, and $b = 1$. The regression situation for genotypes Ab and aB is also shown in the figure. Each of these genotypes gives four points, relating g to e for the four environments, which mark the corners of a parallelogram. The genotypes in fact each gives a regression box rather than a regression line. Such a box cannot usefully be represented by a polynomial regression line—even though of course a cubic line could be found to pass through the four points, it would not be an informative description of the relation between g and e for these genotypes.

The slope of the regression line for genotype AB is $\frac{1}{3}$, and that for genotype ab is $-\frac{1}{3}$. Now the slope of the long sides of the box given by Ab is $\frac{1}{3}$ and that of the short sides is $-\frac{1}{3}$. If, however, we calculate the regression coefficient for this genotype we find it to be $\frac{1}{3}$. This is, of course, because we are minimising the sum of squares of deviation in g . If we minimised the sum of squares of the departures of the four points from the line in a direction normal to the line itself we should again find the slope of the line

FIG. 2.—The regression of g on e where there are two environmental factors α and β and two gene-pairs, A-a responding to α and B-b to β . The regression patterns are shown for genotypes AB and Ab, that for ab being the same as for AB but with the sign changed, and that for aB being the same as for Ab but with the sign changed. In all cases the values of A , a , B and b are chosen to give a straight regression line for genotype AB. Note the box-like form assumed by the regression for genotype Ab. Above: Where $A = 4$, $a = 2$, $B = 2$, $b = 1$. Centre: Where $A = 2$, $a = 1$, $B = 4$, $b = 2$. Below: Where $A = 3$, $a = 1\frac{1}{2}$, $B = 3$, $b = 1\frac{1}{2}$. Note that, as illustrated by the broken line in the top figure, the best fitting regression line for genotype Ab calculated in the customary way does not follow the slope of either side of the regression box. (See in text for further comment.)

to be $\frac{1}{3}$. We might note, too, that if the values of A , a , B , and b had been 2, 1, 4 and 2 respectively, genotype Ab would have given a box with long and short sides interchanged by comparison with fig. 2*a*, as shown in fig. 2 (centre). The regression of g on e would then have worked out at $-\frac{1}{3}$. Thus in a situation such as we are discussing the value of the regression coefficient is dominated by the gene pair having the larger interaction with its relevant environmental factor, the action of the gene pair with the smaller interaction serving only to depress the value of the regression coefficient as found in the customary way. In the extreme case when $A = B$ and $a = b$, the regression box becomes an equal sided parallelogram whose diagonals follow the axes of the graph, and the regression coefficient becomes 0 (fig. 2, bottom).

One further point remains to be made about this simple situation. Genotypes Ab and aB which give regression boxes are the recombinants of AB and ab which give perfect lines. Thus recombination of genotypes which give matching regression lines will necessarily give regression boxes, and recombination of genotypes which give matching boxes will result in them deboxing to give straight lines. We can thus see the basis for Perkins and Jinks (1968*b*) inference that there is a genetical element in the departures from straight regression lines.

6. SETS OF GENES AND PLEIOTROPIC GENES

The simple situation we have been discussing in the previous section requires elaboration in three ways. In the first place there is likely to be not one but a number of genes like A - a , affecting the response to environmental factor α , and similarly a number of genes like B - b affecting the response to β . Furthermore, some of both the A - a and B - b genes will be constant over the genotypes under observation while others will be variable among the genotypes. If we denote by $[A] = A_1 + A_2$ —the algebraic sum of responses of all the relevant genes A_1 , A_2 , etc., in one genotype and by $[a]$ the corresponding sum of all their alleles in an alternative genotype, and similarly denote by $[B]$ and $[b]$ the sums of responses of the B and b genes, we find the condition for a perfectly fitting straight regression line of g on e in genotypes AB and ab to be $[A][b] - [a][B] = 0$. Equally the two genotypes Ab and aB , in which all the A genes have been recombined with all the B genes, gives a regression box unless $[A][B] - [a][b] = 0$. Thus one or other pair of genotypes must give a box unless $[A] = [a] = [B] = [b]$, when there is no detectable genotype $\times$ environment interaction present. Such a perfectly matched set of recombinant lines, in which all A genes had been recombined as a group with all the B genes as a group, must in general be unlikely; but we may observe that virtually any recombination between two genotypes that give a straight regression line will lead to a regression box. Even recombination of one A - a gene with another, or of one B - b gene with another will lead to boxing, for $(A_1 + a_2)/(a_1 + A_2)$ will not generally equal $(A_1 + A_2)/(a_1 + a_2)$, where $A_1 \neq a_1$ and $A_2 \neq a_2$ and if the latter relation fulfils the requirement for a straight regression line the former will not. The exception to this will be where recombination among the B genes has produced new complexes such that $[B]_R/[b]_R = [A]_R/[a]_R$ where the subscript R denotes the recombinant complexes. In such a case where recombination in one set of genes balances in its effect recombination

in the other set, a straight regression line will ensue, but it will not have the same slope as the regression in the original parent genotypes except in the perhaps unlikely event of

$$\frac{[A]}{[a]} = \frac{[B]}{[b]} = \frac{[A]_R}{[a]_R} = \frac{[B]_R}{[b]_R}.$$

When all genes are taken into account, the slope of the regression lines, as noted earlier, will depend on the constant genes as well as the variables, since the constant genes will affect the value of e even though they do not affect the g 's. It might be noted, too, that they can affect the fit of the regression line. Thus, to take an example involving two A and two B genes, a perfectly fitting line will be obtained if

$$(A_1 + A_2)/(a_1 + a_2) = (B_1 + B_2)/(b_1 + b_2) = k,$$

but given this condition, there is no reason to expect that if A_2 is made constant $(A_1 + A_2)/(a_1 + A_2)$ will equal $(B_1 + B_2)/(b_1 + b_2)$. Indeed there is no reason to expect either that $(A_1 + a_2)/(a_1 + a_2)$ or $(B_1 + B_2)/(b_1 + B_2)$ or $(B_1 + b_2)/(b_1 + b_2)$ will have values satisfying the criterion for a perfect fitting regression, though it is of course again possible that a change in the B genes will have an effect balancing that of the A genes so that although the new $[A]/[a]$ does not equal k , it will equal the new $[B]/[b]$ and a straight regression line will again appear, albeit with a different slope. Thus changes in the constancy of the genes can lead to loss of a perfect fitting regression line and the production of a regression box, or of course by a move in the opposite direction the replacement of a box by a line.

Any recombination, or change in the constant genes, is likely to alter a line to a box, whereas only particular recombinations or changes in constancy can lead to the reverse changes from box to line: the majority of recombination or changes of constancy might be expected merely to replace one box by another. The change from box to line might appear therefore to be an event unlikely to be observed. Two further considerations must, however, be borne in mind. In the first place, if there are relatively few genes or closely linked combinations of genes involved, the number of possible recombinant types will be correspondingly low, and if any of these can give a line, lines will be correspondingly likely to appear. Secondly, the observations from which line or box is inferred will be subject to error variation and the test of departure from a straight line will be limited in its sensitivity to the magnitude of the error variation. A box that departs relatively little from a line will then escape detection and will in fact be taken for a line. We are therefore concerned not with the frequency of appearance of perfect fitting lines, but of statistically adequate lines and these will appear much more commonly.

The second elaboration required by our simple situation of the previous section is the recognition that genes must be expected which, unlike the A and B genes we have been considering, respond in their action to changes in both the α and β factors of the environment. Let the response to the α factor of a such a C gene, as we may call it, be C_A and that to the β factor be C_B . Then considering one A gene pair, one B gene pair and one C gene pair the value found for V_e from all eight possible genotypes or from cor-

TABLE 7

Genotypes	V_g	W	Remainder variance
ABC, abc	$\frac{1}{4}[(A-a+C_A-c_A)^2 + (B-b+C_B-c_B)^2]$	$\pm \frac{1}{4}[(A+C_A)^2 - (a+c_A)^2 + (B+C_B)^2 - (b+c_B)^2]$	$\frac{1}{4}[(A+C_A)(b+c_B) - (a+c_A)(B+C_B)]^2/V_e$
ABc, abC	$\frac{1}{4}[(A-a-C_A+c_A)^2 + (B-b-C_B+c_B)^2]$	$\pm \frac{1}{4}[(A+c_A)^2 - (a+C_A)^2 + (B+c_B)^2 - (b+C_B)^2]$	$\frac{1}{4}[(A+c_A)(b+C_B) - (a+C_A)(B+c_B)]^2/V_e$
AbC, aBc	$\frac{1}{4}[(A-a+C_A-c_A)^2 + (B-b-C_B+c_B)^2]$	$\pm \frac{1}{4}[(A+C_A)^2 - (a+c_A)^2 - (B+c_B)^2 + (b+C_B)^2]$	$\frac{1}{4}[(A+C_A)(B+c_B) - (a+c_A)(b+C_B)]^2/V_e$
Abc, aBC	$\frac{1}{4}[(A-a-C_A+c_A)^2 + (B-b+C_B-c_B)^2]$	$\pm \frac{1}{4}[(A+c_A)^2 - (a+C_A)^2 - (B+c_B)^2 + (b+c_B)^2]$	$\frac{1}{4}[(A+c_A)(B+C_B) - (a+C_A)(b+c_B)]^2/V_e$

responding pairs of them (such as ABC and abc) or from any aggregate of such pairs, raised in all four environments will be

$$V_e = \frac{1}{4}[(A+a+C_A+c_A)^2 + (B+b+C_B+c_B)^2]$$

as shown in table 7. The V_g 's, W 's and Remainder Variances are also set out in the table where the genotypes are listed in corresponding pairs, the members of a pair having the same V_g and Remainder Variance but having W 's equal in magnitude but opposite in sign as indicated by a $\pm$ sign in the table.

The pattern is clear, C_A and c_A go with either A or a , according to whether C is associated with A and c with a , or vice versa, in the genotypes. Similarly, C_B and c_B go with B or b according to the association of $B-b$ and $C-c$ in the genotypes. Thus C_A and c_A act as parts of $[A]$ and $[a]$ and C_B and c_B as parts of $[B]$ and $[b]$, the only special feature being that C_A can never recombine with C_B nor c_A with c_B . C genes may therefore be regarded as no more than combinations of A and B genes with a restriction on their recombination. The effect of such recombination as does occur, whether between $A-a$ and $B-b$ or between either of them and $C-c$ is as before: it will change a line to a box or sometimes vice versa. One special point requires notice. If C_A and c_A bear the same ratio to A and a as C_B and c_B do to B and b , recombination of ABC and abc to give ABc and abC will not destroy any straight regression line given by ABC and abc; ABc and abC will also give a straight line though of course with a different slope (fig. 3). Thus the introduction of C genes, with their pleiotropic action, into the system brings no new principle into the relation between genotype, regression line and regression box, though it will place restrictions on the frequencies with which one is changed into the other and on the detailed appearance of the changes.

7. INCOMPLETE SETS OF GENOTYPES AND ENVIRONMENTS

So far we have assumed that all the genotypes representing the possible combinations of the genes under discussion, and all the environments representing combinations of the factors under discussion, are equally present. Thus with $A-a$ and $B-b$ we have brought all genotypes, AB, Ab, aB and ab into account, and with the environmental factors α and β , all the four environments, $\alpha_1\beta_1$, $\alpha_1\beta_2$, $\alpha_2\beta_1$ and $\alpha_2\beta_2$. This is achievable in at least some experiments, which are the more informative in consequence; but we can hardly assume it to be the case in natural populations (or even in the true-breeding lines derived from them) or in a complex range of natural environments. Our third elaboration must therefore be to consider the properties of observations on incomplete sets of genotypes and environments. Consideration will be restricted to the variants of the two-gene two-environmental factor situation with which we started.

The only reduction we need to take into account will be from four to three genotypes and from four to three environments, since a reduction to two would amount to discussing a single gene system (though the contrast might be between $A+B$ and $a+b$ or between $A+b$ and $a+B$) or a single environmental factor (which might be $\alpha_1+\beta_1$ versus $\alpha_2+\beta_2$ or $\alpha_1+\beta_2$ versus $\alpha_2+\beta_1$). Now, whether for genotypes or environments, the missing type might be one of the extremes, AB or ab for genotypes and $\alpha_1\beta_1$ or

$\alpha_2\beta_2$ for environments, or it might be one of the central types Ab or aB for genotypes and $\alpha_1\beta_2$ or $\alpha_2\beta_1$ for environments. The omission of an extreme type will be referred to as giving symmetrical sets, since the set of genotypes AB , Ab and aB , which are left when we omit ab , are symmetrical in respect of a and b , as environments $\alpha_1\beta_1$, $\alpha_1\beta_2$ and $\alpha_2\beta_1$, left when $\alpha_2\beta_2$ is omitted,

FIG. 3.—The regression of g on e where there are two environmental factors α and β , and three gene-pairs, A - a responding to α , B - b responding to β and C - c responding to α and β simultaneously. The regressions are shown for four of the eight genotypes, ABC and Abc above and ABc and AbC below. The regression for the remaining four genotypes are the same as for the four shown, but with the signs reversed. The values assumed are $A = 8$, $a = 4$, $B = 4$, $b = 2$, $C_A = 4$, $C_A = 2$, $C_B = 2$, $c_B = 1$. Note that both genotypes ABC and ABc give straight regression lines, albeit of different slopes. The regression box for genotype AbC is related to the line for ABC and that for Abc is related to the line for ABc .

are in respect of α_2 and β_2 . The omission of a central genotype or environment will be referred to as leaving an asymmetrical set, since, for example, AB, Ab and ab are not symmetrical in a and b, nor $\alpha_1\beta_1$, $\alpha_2\beta_1$ and $\alpha_2\beta_2$ in respect of α_2 and β_2 . There will thus be two cases to consider when we omit a genotype (though they can be formally converted into one another by interchanging B and b) and two cases when we omit an environment (though again they are formally interconvertible by interchange of β_1 and β_2). Despite this interconvertibility, however, it is instructive to set out the results from symmetrical and asymmetrical sets separately. In addition to the situation with three genotypes in all four environments and all four genotypes in three environments, we will also consider the four possible cases of three genotypes in three environments, the four being symmetrical/symmetrical, symmetrical/asymmetrical, asymmetrical/symmetrical and asymmetrical/asymmetrical, the set of genotypes being described before the line and the set of environments after it. We shall omit ab to give a symmetrical set of genotypes and aB to give an asymmetrical set, and similarly omit $\alpha_2\beta_2$ and $\alpha_1\beta_2$ to give symmetrical and asymmetrical sets of environments.

The expressions for V_e , V_g , W and Remainder Variance are shown for all these cases, together with the basic case of all four genotypes in all four environments in table 6. With four genotypes the expressions are given for these in pairs, since AB and ab always have the same values for V_g and Remainder Variance and values of W equal except for taking opposite signs (as shown by $\pm$ in the table) as do Ab and aB. But with only three genotypes the expressions are given for all of them since although if given for AB and Ab those for AB or ab are derivable, they are perhaps not always immediately obvious. The regressions of g on e are also shown in fig. 4 for the case where $A = 4$, $a = B = 2$, and $b = 1$, which we have of course already seen to give a straight line where all four genotypes are grown in all four environments (fig. 2).

The most striking feature of both table 6 and fig. 4 is the robustness of the linear relation. Where $Ab - aB = 0$ and AB and ab consequently give a linear regression, the omission of an environment, whether central or extreme, leaves the linear regression unchanged, even the value of the regression coefficient being unaffected. The omission of an extreme genotype, to give a symmetrical residual set, also leaves the regression of AB (or ab if it is the extreme genotype retained) still linear though the value of the regression coefficient is reduced. The omission of a central genotype, to give an asymmetrical residual set, does destroy the linearity of the AB regression (where aB is the genotype omitted). The ab regression also is no longer perfectly linear, but it does not depart in any way substantially from linearity. The same is true when both a genotype and an environment are missing. The AB regression is still perfectly linear when the residual genotypes constitute a symmetrical set, and where a central genotype is omitted to give an asymmetrical set the ab regression, though not perfectly linear is very nearly so. Indeed, when we take $A = 4$, $a = B = 2$, and $b = 1$, the best fitting straight line found for the ab regression never fails to take up at least 95 per cent of the variation in g , whether all four environments are present or, if one is omitted, whether this be a central or an extreme environment. In the great majority of experiments such a small departure from strict linearity would pass undetected as significant. Thus

at least one of the linear regressions is retained either exactly or to a close approximation even if a genotype or an environment is missing.

With only three environments, there can be only three points on the g/e regression, which thus cannot give a four-sided box of the kind we discussed

FIG. 4.—The regression of g on e for the same model as in fig. 2 (top), but with one genotype missing, one environment missing or both one genotype and one environment missing. Note that the absence of an environment never destroys the linear regression of genotype AB. The absence of a genotype also leaves this linear regression intact in half the cases. In the other half the regression for AB loses its linearity but that for genotype ab remains almost linear. For the meaning of Sym and Asym, see in the text. Where there are four genotypes only two of the regressions are shown (as in fig. 2), but with three genotypes all three regressions are shown. With four environments the quadrilateral regression boxes are shown complete. With three environments the boxes should be triangular, but in all cases the long side of the triangle has been omitted for clarity. All the regressions are on the same scale, which is marked on the regression at the top left.

in earlier sections. The triangular box that is left is much less revealing of its true nature and could easily pass for a quadratic regression relation. Since, however, we can now recognise the triangle as the residuum of a quadrilateral box, we can see that fitting a quadratic regression would not be particularly useful.

One last point remains to be made about the simultaneous omission of genotype and environment. It will be seen from fig. 4 that with symmetrical residual sets of both genotypes and environments the triangular boxes given by genotypes Ab and aB depart widely from straight lines. With an asymmetrical residual set of genotypes, and even more with such a set of environments, the triangular boxes are drawn out and flattened, so approaching more nearly to a straight line. With both genotypes and environments in asymmetrical sets the triangles are very much drawn out and flattened. Indeed, with the error variation experienced in many experiments all three regressions and not just that of genotype ab, might well fail to show a significant departure from linearity. Thus, far from destroying the evidence of linearity of the extreme genotypes in this example, the omission of a genotype and particularly of an environment can result in a spurious approach to linearity by the regression of those genotypes which are not expected to show it. The linear regressions found in experiments using natural environments may well owe something of their appearance to this phenomenon. Clearly the whole subject of incomplete sets of genotypes and environments would well repay further investigation.

Acknowledgments.—I am happy to acknowledge financial support from the Leverhulme Trust and the Agricultural Research Council. I am indebted to both Professor J. L. Jinks, F.R.S., and Dr P. D. S. Caligari for very valuable discussion of points that arose during this work.

8. REFERENCES

- FREEMAN, G. H., AND PERKINS, J. M. 1971. Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. *Heredity*, 27, 15-23.
- FRIPP, Y. J. 1972. Genotype-environmental interactions in *Schizophyllum commune*. II. Assessing the environment. *Heredity*, 28, 223-238.
- FRIPP, Y. J., AND CATEN, C. E. 1971. Genotype-environmental interactions in *Schizophyllum commune*. I. Analysis and character. *Heredity*, 27, 393-407.
- MATHER, K., AND CALIGARI, P. D. S. 1974. Genotype $\times$ environment interactions. 1. Regression of interaction on overall effect of the environment. *Heredity*, 33, 43-59.
- MATHER, K., AND JONES, R. M. 1958. Interaction of genotype and environment in continuous variation. I. Description. *Biometrics*, 14, 343-359.
- PERKINS, J. M., AND JINKS, J. L. 1968a. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity*, 23, 339-356.
- PERKINS, J. M., AND JINKS, J. L. 1968b. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity*, 23, 525-535.