

THE ANALOGY BETWEEN FACTORIAL EXPERIMENTATION AND BALANCED MULTI-POINT LINKAGE TESTS

W. F. BODMER and P. A. PARSONS

Department of Genetics, University of Cambridge

Received 5.vi.58

1. INTRODUCTION

It was pointed out by Fisher (1952) in the Bateson lecture that the factorial method of experimentation, now used extensively in agriculture and other fields of research, derives its name and structure from the simultaneous segregation of Mendelian characters.

Kempthorne (1955) stressed the genetic origins of the ideas of factorial experimentation when he used these techniques in the splitting up of the genetic variance into components in a system involving the average effects and interactions of unspecified polygenes. In this paper, however, we shall consider major gene effects and their interactions such as are available from a multiple linkage experiment.

In recent years multiple linkage data have been analysed with series of simple χ^2 tests. A discussion of these has been given by Fisher (1949), Parsons (1957, 1958) and Wallace (1947, 1957). In this paper, by applying the principles of factorial analysis, we are attempting to give an alternative and rather more integrated method of analysis.

2. THE FACTORIAL SCHEME

In the data obtained from a backcross linkage test involving the three factors *A*, *B*, *C* and their recessive alleles, each factor will be either heterozygous *Aa*, *Bb*, *Cc* or homozygous *aa*, *bb*, *cc*. This is directly analogous to three treatments at two levels in a 2^3 factorial experiment in which there are eight possible treatments which can be represented as 1, *a*, *b*, *c*, *ab*, *ac*, *bc*, *abc*. The letters *a*, *b*, *c* denote one of the treatment levels and may be taken to represent the recessive homozygotes *aa*, *bb*, *cc* in the genetical analogy. Thus the treatment effects of the factorial experiment correspond to the genotypic viabilities. The main effect in the factorial experiment corresponds to the main viability effect of a gene and interactions between main effects to viability interactions.

The blocks in the factorial experiment correspond to the modes of gamete formation. If each genotype is represented equally in each mode of gamete formation (or block), we can carry out an analysis of variance in the usual way, and hence obtain an analysis of χ^2 , giving measures of the viability effects and interactions, and of the recombination effect (block effect). This implies that all the possible multiple heterozygotes must be used as parents, and that each must

contribute equally to the data. In this case the experiment may be called "orthogonally balanced". A sum of squares (S.S.), and hence a χ^2 for the differences in parental heterozygote contributions can be calculated which is a measure of the departure from orthogonality. Yates and Cochran (1938) have shown in similar situations that such a S.S. may be adequate to account for the non-orthogonality provided that deviations from orthogonality are not too severe. If there are severe deviations, a weighted analysis ought to be used. In sections 3 and 4 a full analysis of χ^2 for data from three-point backcross experiments will be described.

As in agricultural factorial experiments, the higher-order viability interactions may, intuitively, be expected to be of no importance. Even if information were lacking on some of the higher-order interactions it is, nevertheless, probable that the estimates of the recombination fractions would not in fact be biased. In agricultural factorial experiments information on higher-order interactions is often sacrificed so as to be able to use smaller blocks, this process being termed confounding (Fisher, 1935-53), and the same may be done in a genetical linkage experiment.

For example, in a four-point test, the eight possible quadruple heterozygotes can be divided into two sets of four by confounding the $abcd$ or 4-factor interaction where d represents the fourth factor. This interaction can be found by evaluating $(a-1)(b-1)(c-1)(d-1)$ which equals:

$$1 - a - b - c - d + ab + ac + ad + bc + bd + cd - abc - abd - acd - bcd + abcd$$

Confounding this interaction means that the terms of this expansion with negative signs (1 and 3 factor effects) are in one block, and those with positive (0, 2, and 4 factor effects) in the other. Considering those with negative sign, it can be seen that a, b, c, d represent the same quadruple heterozygotes as bcd, acd, abd, abc so that one set would be:—

$$\frac{aBCD}{Abcd}, \frac{AbCD}{aBcd}, \frac{ABcD}{abCd}, \frac{ABCD}{abcd}$$

and the other

$$\frac{ABcd}{abCD}, \frac{AbCd}{aBCD}, \frac{AbcD}{aBCd}, \frac{ABCD}{abcd}$$

Hence it is essential to assume that the sum of the 0, 2 and 4 factor effects is equal to the sum of the 1 and 3 factor effects in estimating recombination from one such set in a four-point linkage test. If this assumption can be made, and there is orthogonality within a set, it can easily be seen that recombination may always be estimated by the addition of the numbers in each mode of gamete formation without risk of bias.

When there are both coupling and repulsion data for every possible pair of factors in a multi-point linkage test, the resulting set of data is

said to be balanced (Fisher and Mather, 1936; Wallace, 1957). In a two-point linkage test, the two possible double heterozygotes are necessary for balance in this sense, and in the three-point test the four possible triple heterozygotes. However, in the four-point test, each of the two sets of four defined above by confounding the four-factor interaction is balanced. Such a set of four may be called a balanced set.

This condition for balance corresponds to the condition that no two-factor interactions should be confounded. The nature of the occurrence of the genotypes in complementary pairs ensures that the

TABLE 1a
Genotypes in a 4×4 Latin square

Parental heterozygote	Mode of gamete formation			
	(0)	(1)	(2)	(12)
$abc/+++$	abc I	a bc	c ab	b ac
$+bc/a++$	a bc	abc I	b ac	c ab
$ab+/+b+$	c ab	b ac	abc I	a bc
$a+c/+b+$	b ac	c ab	a bc	abc I

main effects are never confounded and so only one parental heterozygote is needed if all the interactions are negligible.

In five- and higher-point data, the problem of selecting balanced sets is complicated and may involve confounding several interactions. A scheme in which balanced sets may be found for higher-order experiments has been devised by Edwards (1958).

3. THE THEORY OF THE χ^2 ANALYSIS

The data from a balanced three-point backcross linkage experiment can be set out as in table 1a (*cf.* Parsons, 1957), where a genotype is designated by small letters for the loci at which it is homozygous. The expectations assumed for each observed class are usually taken to be the product of three quantities representing respectively a contribution from the parental heterozygote, a contribution from the mode of gamete formation, and a contribution for the viability of the particular genotype observed. The expectations for the observed classes, corresponding to the genotypes listed in table 1a, are given in table 1b,

where allowance is made for possible interactions between these three sources of variation. The parameters are defined as follows:—

- a_i = proportionate contribution from i^{th} parental heterozygote ;
- p_j = proportionate contribution from j^{th} mode gamete formation ;
- v_k = proportionate contribution from the k^{th} genotype ;
- e_{ijk} = proportionate contribution for the interaction of the k^{th} genotype with the i^{th} heterozygote and the j^{th} mode of gamete formation.

There are thirty-two observed classes, leaving thirty-one degrees of

TABLE 1b
Expected proportions corresponding to the genotypes in 1a

Parental heterozygote	Mode of gamete formation			
	(0)	(1)	(2)	(12)
$abc +++$	$a_1 p_1 v_1 e_{111}$ $a_1 p_1 v_2 e_{112}$	$a_1 p_2 v_3 e_{123}$ $a_1 p_2 v_4 e_{124}$	$a_1 p_3 v_5 e_{135}$ $a_1 p_3 v_6 e_{136}$	$a_1 p_4 v_7 e_{147}$ $a_1 p_4 v_8 e_{148}$
$+bc a++$	$a_2 p_1 v_3 e_{213}$ $a_2 p_1 v_4 e_{214}$	$a_2 p_2 v_1 e_{221}$ $a_2 p_2 v_2 e_{222}$	$a_2 p_3 v_7 e_{237}$ $a_2 p_3 v_8 e_{238}$	$a_2 p_4 v_5 e_{245}$ $a_2 p_4 v_6 e_{246}$
$ab+ ++c$	$a_3 p_1 v_5 e_{315}$ $a_3 p_1 v_6 e_{316}$	$a_3 p_2 v_7 e_{327}$ $a_3 p_2 v_8 e_{328}$	$a_3 p_3 v_1 e_{331}$ $a_3 p_3 v_2 e_{332}$	$a_3 p_4 v_3 e_{343}$ $a_3 p_4 v_4 e_{344}$
$a+c +b+$	$a_4 p_1 v_7 e_{417}$ $a_4 p_1 v_8 e_{418}$	$a_4 p_2 v_5 e_{425}$ $a_4 p_2 v_6 e_{426}$	$a_4 p_3 v_3 e_{433}$ $a_4 p_3 v_4 e_{434}$	$a_4 p_4 v_1 e_{441}$ $a_4 p_4 v_2 e_{442}$

freedom after the restriction that the total observed must equal the total expected has been taken into account.

Let us assume that

$$a_i = \frac{1}{4}(1 + a_i), p_j = \frac{1}{4}(1 + \pi_j), v_k = \frac{1}{2}(1 + \lambda_k) \text{ and } e_{ijk} = 1 + \mu_{ijk}.$$

Then the expectation of the observed total from the i^{th} heterozygote, in the j^{th} mode of gamete formation for the k^{th} genotype is

$$m(1 + a_i + \pi_j + \lambda_k + \mu_{ijk} + \text{higher-order terms})$$

if it is assumed that a_i, π_j, λ_k and μ_{ijk} are small and $m = N/32$ where N is the observed total. Thus, when the deviations of the proportionate contributions from zero are small, the expectations are approximately linear functions of the parameters.

We have defined four parameters each for the heterozygote and mode of gamete formation effects, and eight parameters for the viabilities or genotype effects. These must be subject to the restrictions

$$\sum_i a_i = 0, \sum_j \pi_j = 0, \sum_k \lambda_k = 0 \quad . \quad . \quad (3.1)$$

leaving thirteen independent parameters for these effects. The thirty-two parameters μ_{ijk} are subject to the restrictions

$$\sum_{\text{given } i} \mu_{ijk} = 0, \quad \sum_{\text{given } j} \mu_{ijk} = 0, \quad \sum_{\text{given } k} \mu_{ijk} = 0, \quad \sum_{\text{all } ijk} \mu_{ijk} = 0 \quad (3.2)$$

of which fourteen are independent. This leaves eighteen independent parameters to describe interaction between genotypes, heterozygotes and modes of gamete formation which, together with the thirteen parameters describing the effects discussed above, make up the thirty-one available degrees of freedom.

If values of the parameters are fitted assuming that $e_{ijk} = 1$ for all i, j and k by the method of maximum likelihood, the equations for estimation of the remaining thirteen parameters are given by equating observed row, column and diagonal totals to their expected values. These imply in fact the restrictions 3.2. It is also clear that assuming separate viabilities for the pairs of genotypes in a cell does not affect the estimation of the relative frequencies of the modes of gamete formation.

Now the general procedure followed in applying Fisher's method of scoring (*cf.* Fisher 1956) is as follows. Suppose we make observations on n classes, and the expectation of the r^{th} observation a_r is $M_r(\theta_1 \dots \theta_k)$, where $\theta_1 \dots \theta_k$ are parameters to be estimated. Then we consider the k linear scores

$$S_i = \sum_{r=1}^n a_r \frac{\partial}{\partial \theta_i} (\log M_r) = \sum_{r=1}^n K_{ir} a_r \text{ say}$$

corresponding to the k parameters, and also the information matrix $I_{ij} = \sum_{r=1}^n K_{ir} K_{jr} M_r$. If now m_r and k_{ir} are the values taken by M_r and K_{ir} when, say, $\theta_1 = 0 = \theta_2 = \dots = \theta_k$, approximations to the maximum likelihood estimates of $\theta_1 \dots \theta_k$ are given by the equations $\sum_{j=1}^k I_{ij} \theta_j = S_i$.

If the matrix I_{ij} is diagonal, that is $I_{ij} = 0$ when $i \neq j$, the parameters θ_i may be termed orthogonal and the approximate estimates of them are $\theta_i = S_i/I_i$ where $I_i = I_{ii}$. An approximate test of significance for the departure of θ_i from zero is then given by $\chi^2 = S_i^2/I_i$. More accurate estimates and tests are obtained by repeating the scoring procedure with the values of θ_i given above instead of taking all the θ_i to be zero. Now if we assume the state $\theta_1 = 0 = \dots = \theta_k$ corresponds to equality of the expected classes, we have

$$M_r = m[1 + k_{ir}\theta_i + o(\theta^2)]$$

where k_{ir} has the meaning attached to it above and $m = N/n$. Thus k_{ir} give the scores appropriate for a linear system, whatever the values of the θ_i and so scoring at $\theta_i = 0$ is equivalent to assuming a linear system. This is the assumption made in the analysis of variance, where the various effects are made to form a set of orthogonal parameters and

the total sum of squares (S.S.) is split into components which measure the extent of these effects. Now the χ^2 testing departure from equality of the observations a_r is $\chi_{n-1}^2 = \sum \frac{(a_r - N/n)^2}{N/n} = \sum \frac{a_r^2}{m} - N = \frac{1}{m} \times \text{S.S.}$ of the a_r . If, therefore, we define the parameters θ_i corresponding to the various effects as for an analysis of variance, we have

$$S_i = \sum_{i=1}^n k_{ir} a_r, I_i = \sum_{r=1}^n k_{ir}^2 m$$

and the Total S.S. = $\sum (a_r - m)^2 = m \sum_{i=1}^k S_i^2 / I_i + \text{Residual S.S.}$ with $n-1-k$ degrees of freedom where $m S_i^2 / I_i$ is the component of the S.S. for the i^{th} effect, and on division by m this provides the χ^2 detecting the departure of θ_i from zero.

Hence if we carry out an analysis of variance on the data from such a linkage experiment, and divide the component S.S. by m , we obtain an exactly analogous partition of χ^2 for detecting departures from equality of the expected classes. The accuracy of the χ^2 tests thus obtained on the assumed expectations given in table 1*b* depends on the departures from equality being small, so that the expectations are to a good approximation linear functions of the parameters.

The total S.S. for the observations from such a three-point backcross linkage experiment is

$$\begin{aligned} & \sum y_{ij1}^2 + \sum y_{ij2}^2 - 2n \left[\frac{\sum y_{ij1} + \sum y_{ij2}}{2n} \right]^2 \\ &= \frac{1}{2} \left[\sum (y_{ij1} + y_{ij2})^2 - n \left(\frac{\sum (y_{ij1} + y_{ij2})}{n} \right)^2 \right] \\ &+ \frac{1}{2} \left[\sum (y_{ij1} - y_{ij2})^2 - n \left(\frac{\sum (y_{ij1} - y_{ij2})}{n} \right)^2 \right] + \frac{n}{2} \left[\frac{\sum (y_{ij1} - y_{ij2})}{n} \right]^2 \end{aligned}$$

where y_{ij1} and y_{ij2} are the two observed totals for the complementary genotypes from the i^{th} heterozygote and the j^{th} mode of gamete formation, and $n = 16$. In other words the total S.S. consists of $\frac{1}{2}$ [S.S. of the sums of complementary genotypes] + $\frac{1}{2}$ [S.S. of the differences of complementary genotypes] + $\frac{1}{2}$ S.S. due to $\sum (y_{ij1} - y_{ij2})$. The sums and differences of the complementary genotypes form two 4×4 Latin squares and the S.S. from $\sum (y_{ij1} - y_{ij2})$ is the S.S. for the total of the differences square. The "sums" square is the form in which the data are usually presented for the detection of linkage and interference (see Fisher (1949) and Wallace (1957)). The S.S. for the Latin squares can be partitioned in the usual way, namely into components for rows, columns, diagonals and an error term.

If the α_i as defined above are all equal, then the experiment is orthogonally balanced, that is, all the different genotypes are represented equally for each mode of gamete formation. In this case, the

usual treatment effects and interactions represent the corresponding viability effects and interactions of the various loci, as described in the previous section. For example, the S.S. for the two-factor interaction is, symbolically,

$$4 \times \frac{1}{8} [\{(a-1)(b-1)(c+1)\}^2 + \{(a-1)(b+1)(c-1)\}^2 + \{(a+1)(b-1)(c-1)\}^2]$$

i.e. $4 \times \frac{1}{8} \left[\begin{array}{l} \{(abc+1) + (c+ab) - (a+bc) - (b+ac)\}^2 \\ + \{(abc+1) + (b+ac) - (c+ab) - (a+bc)\}^2 \\ + \{(abc+1) + (a+bc) - (b+ac) - (c+ab)\}^2 \end{array} \right]$

$$= 4 \times \frac{1}{8} \left\{ 4 \left[\begin{array}{l} (abc+1)^2 \\ + (a+bc)^2 \\ + (b+ac)^2 \\ + (c+ab)^2 \end{array} \right] - \left[\begin{array}{l} (abc+1)^2 \\ + (a+bc)^2 \\ + (b+ac)^2 \\ + (c+ab)^2 \end{array} \right]^2 \right\}$$

where *abc* etc. represent the means of the contributions of the corresponding genotypes to the four modes of gamete formation. This is one-half the diagonals S.S. for the "sums" square. In a similar way it can be shown that the diagonals S.S. for the "differences" square is twice that for the main viability effects, and the S.S. for the total of the "differences" square is the S.S. for the three-factor interaction. The expectation of each term in the "differences" square consists only of the difference of the viability contributions from complementary genotypes if the interactions μ_{ijk} are all zero. Hence the S.S. for rows in the "differences" square measures the interaction of viabilities and heterozygotes, and the S.S. for columns, the interactions of viabilities and modes of gamete formation. We therefore obtain the scheme for the analysis of variance given in table 2.

TABLE 2
Scheme for analysis of variance

	d.f.
$\frac{1}{2}$ "sums" Latin square	
Parental Heterozygote (rows)	3
Recombination (columns)	3
2-factor interactions (diagonals)	3
Error	6
	—
	15
$\frac{1}{2}$ "differences" Latin square	
Viability \times Parental heterozygote (rows)	3
Viability \times Recombination (columns)	3
Main viability effects (diagonals)	3
Error	6
	—
	15
3-factor interaction	
(S.S. from total of "differences" squares)	1
	—
Total	31

There are in fact 12 degrees of freedom for error, so that only 6 of the 18 degrees of freedom available for describing interactions between the genotypes, the heterozygotes and the modes of gamete formation have been used. These 6 d.f. are made up of 3 d.f. each for the viability \times heterozygote and viability \times recombination interactions. The remaining unused 12 d.f. would be used to account for recombination \times heterozygote and other higher-order interactions. High components for the two "error" terms would indicate the existence of such disturbances.

On division by $m = N/32$ the various components of the analysis of variance will be turned into χ^2 s for the corresponding degrees of freedom.

Excluding maternal effects, a viability \times heterozygote interaction would indicate a "competition" effect, that is a dependence of the viability of a genotype on the relative proportions of the genotypes with which it is competing. Such "competition" effects may be expected to occur quite generally and to be of considerable importance in the study of population genetics. Haldane (1924) has investigated one form of such an effect occurring in populations, which he has termed "familial selection". The competition effect would be indicated by the viability \times heterozygote interaction, because the different genotypes are represented in differing proportions amongst the offspring of each heterozygote. A balanced linkage experiment may thus be of importance when regarded as a means for detecting viability competition effects due to a number of linked loci.

Recombination \times heterozygote interactions are rare, though Parsons (1957) has observed a variation in recombination according to the parental heterozygote which is such an interaction. The occurrence of viability \times recombination interactions is most unlikely, as such an effect is almost inconceivable. The interpretation of terms from the "differences" square will be unaltered by significant non-orthogonality as these depend only on the difference of the complementary genotypes, for which the heterozygote contributions must necessarily be equal.

It is clear that a similar analysis of χ^2 can be obtained for any balanced two- or higher-point experiments.

The accuracy of the analysis presented above, which is based on the multiplicative expectations given in table 1*b*, depends on being able to take the expectations as approximately linear expressions in the parameters which is only possible if the effects are small. If, however, logarithms are taken, the expectations are then linear functions of the logarithms of the various parameters. The scores for the logarithms of the parameters are therefore constant, and the information may be expected to vary little with variation in the parameters. It would seem, therefore, that a more accurate analysis could be obtained by doing the calculations on the logarithms of the observed data.

By comparing the analysis obtained before and after taking logarithms it is possible in some cases to distinguish between additive

and multiplicative action of genes. This subject is considered further by Bodmer (1959).

4. APPLICATIONS OF THE ANALYSIS OF χ^2 TO PUBLISHED DATA ON THREE-POINT EXPERIMENTS

(i) Data of Wallace (1947)

The data are for a balanced three-point test involving the factors *wavy-2* (*wv2*), *shaker-2* (*sh2*) and sex in the house mouse (*Mus musculus*). The analysis of χ^2 is presented in table 3.

TABLE 3
Analysis of χ^2 for Wallace's (1947) data

Latin square of sums	d.f.	S.S.	χ^2	P in %
Rows (parental heterozygotes)	3	57.84	4.09	30 - 20
Columns (recombination)	3	1124.59	79.44	$\ll 0.1$
Diagonals (two-factor interactions)	3	5.09	0.36	95 - 90
Error	6	116.19	8.21	30 - 20
	15	1303.71	92.10	$\ll 0.1$
Latin square of differences				
Rows (viability \times parental heterozygote)	3	62.59	4.42	30 - 20
Columns (viability \times recombination)	3	19.34	1.37	80 - 70
Diagonals (main viability effects)	3	1.09	0.08	>99
Error	6	158.69	11.21	10 - 5
	15	241.71	17.08	
Three-factor interaction	1	42.78	3.02	10 - 5
	31	1588.20	112.20	$\ll 0.1$

The overall departure from equality of the observed classes is clearly highly significant. By far the largest component is that for recombination, which is, in fact, the only significant component. However, the χ^2 s for the two-factor interaction and the main effects are suspiciously small compared with that for the three-factor interaction. The error χ^2_6 in the differences square is also on the verge of significance at 5 per cent. indicating the possible existence of some disturbance, not taken into account in the analysis, as discussed in the previous section. It is clear that, as shown in Wallace's original paper, there is no evidence for non-orthogonality or two-factor viability interactions which might disturb the crude estimates of recombination. The χ^2_3 obtained by Fisher (1949) for the two-factor

interactions by more complicated analysis was 0.43, leaving a deviation χ^2_6 of 7.40, which corresponds with the "error" term of the sums square. These agree well with the results presented in table 3. There is no evidence from the data of any viability \times heterozygote interaction or competition effect and, as expected, there is no viability \times recombination interaction.

TABLE 4
Analysis of χ^2 for Parsons' (1957) data

Latin square of sums	d.f.	S.S.	χ^2	P in %
Parental heterozygotes	3	14,822	50.3	$\ll 0.1$
Recombination	3	2,763,091	9,374.4	$\ll 0.1$
Two-factor interactions	3	2,056	7.0	10 - 5
<i>ab</i> -interaction	1	435	1.5	30 - 20
<i>ac</i> -interaction	1	780	2.6	20 - 10
<i>bc</i> -interaction	1	841	2.9	10 - 5
Error	6	24,781	84.0	$\ll 0.1$
Total	15	2,804,750	9,515.7	
Latin square of differences				
Viability \times parental heterozygote	3	2,121	7.2	10 - 5
Viability \times recombination	3	76	0.26	98 - 95
Main viability effects	3	18,928	64.2	$\ll 0.1$
main effect of <i>a</i>	1	648	2.2	20 - 10
main effect of <i>b</i>	1	17,955	60.9	$\ll 0.1$
main effect of <i>c</i>	1	325	1.1	30 - 20
Error	6	22,997	78.0	$\ll 0.1$
Total	15	44,122	149.66	
Three-factor interaction (<i>abc</i>)	1	0		
Total	31	2,848,872	9,665.36	$\ll 0.1$

(ii) *Data of Parsons (1957)*

These data are for the three loci *scute* (*sc*), *crossveinless* (*cv*) and *vermilion* (*v*) situated on the sex chromosome of *Drosophila melanogaster*. The analysis of χ^2 is presented in table 4.

The departure from equality is again wildly significant, with the main contribution from the recombination effect. As predicted in the last section, the variation in recombination according to parental heterozygote found by Parsons, is indicated by highly significant error terms in both squares. There is a significant departure from orthogonality, as might be expected when the numbers are so large, but this is not severe. The only viability effect is the main effect due to crossveinless, which is highly significant. There is no evidence for any other interactions, although the *bc*-interactions are a little inflated.

The value of χ_3^2 for the two-factor interaction obtained by Parsons (1957) was 5.9 by the usual crude method and 12.6 by Fisher's (1949) method. The value obtained above lies between these two. The residual χ_6^2 of 41.4 given by Parsons is somewhat less than the error terms of either of the Latin squares. However, with such large effects as those indicated for Parsons' data, the theory of the last section would suggest some inaccuracy if scoring is done on the assumption of equal expectations.

Both of these analyses therefore show that the mode of analysis employed in agricultural factorial experimentation gives results in agreement with the analyses using simple χ^2 tests. Some deviation from orthogonality can be tolerated without serious bias as in the *Drosophila* data. Furthermore, this method of calculation has the advantage that all major sources of variation can be detected in one computation.

5. DISCUSSION

The method of analysis of multiple linkage data discussed above enables one to detect and isolate the major sources of variation without much difficulty. The usual model assumed for the estimation of recombination fractions implies no competition effects, and in the case of non-orthogonality may involve heavy computations.

Viability interactions in an orthogonal experiment do not affect estimation of recombination by the addition of the column totals of a Latin square, as they will cancel out (sections 2 and 3). For moderate deviations from orthogonality, with small interactions, as usually occur in practice, addition of the column totals will not result in much error (Parsons, 1957). It seems probable, therefore, that errors in the model which allow for genotypic viability differences are likely to be of the same order as the error in the crude method of estimation obtained by the addition of column totals. Each case must, however, be judged on its own merits.

From the practical point of view, when estimates of recombination are required, it is necessary to do a balanced experiment so allowing for these disturbances. It is clear that deviations from orthogonality should not be too severe.

The analysis of χ^2 provides useful information on viabilities and the form of their interactions associated with major genes which may possibly have important applications in population genetics. A similar method of analysis can be used for other sources of variation associated with major genes. An example would be the fertility and vigour associated with segregating blood-group genes.

6. SUMMARY

1. By applying the principles of the analysis of factorial experiments in agriculture, a comprehensive method of analysing multiple balanced backcross linkage experiments is presented.

2. The technique of confounding provides an interpretation of balance in terms of confounded interactions, and shows what a balanced set as used in multi-point linkage tests implies.

3. It is possible, using this method of analysis, to obtain measures of all the major sources of variation such as the viability effects and interactions, recombination, and competition effects. The implications of performing an analysis on the logarithms of the observed data are briefly discussed.

4. The method is illustrated on data from two balanced three-point experiments.

5. The implication of the method in the estimation of recombination and other possible applications are discussed.

Acknowledgment.—We would like to thank Dr A. R. G. Owen for helpful discussions in the course of this investigation. One of us (W. F. B.) wishes to acknowledge receipt of an Agricultural Research Council Studentship.

7. REFERENCES

- BODMER, W. F. 1959. Multiplicative effects and the logarithmic transformation in the analysis of balanced multi-point linkage tests. *Heredity*, 13, 157-164.
- EDWARDS, A. W. F. 1958. Number of mating types required for balance in multi-point linkage programmes. *Nature*, 181, 503-504.
- FISHER, R. A. 1935-53. *The Design of Experiments*. Oliver and Boyd, Edinburgh.
- FISHER, R. A. 1949. Note on the test of significance for differential viability in frequency data from a complete three-point test. *Heredity*, 3, 215-219.
- FISHER, R. A. 1952. Statistical methods in genetics. *Heredity*, 6, 1-12.
- FISHER, R. A. 1956. *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh.
- FISHER, R. A., AND MATHER, K. 1936. A linkage test with mice. *Ann. Eug.*, 7, 265-280.
- HALDANE, J. B. S. 1924. A mathematical theory of natural and artificial selection. *Trans. Cambridge Phil. Soc.*, 23, 19-41.
- KEMPTHORNE, O. 1955. The correlation between relatives in random mating populations. *Cold Spr. Hbr. Symp. Quant. Biol.*, 20, 60-78.
- PARSONS, P. A. 1957. An effect of gene arrangement on the recombination fraction of *Drosophila melanogaster*. *Heredity*, 11, 117-127.
- PARSONS, P. A. 1958. A balanced four-point linkage experiment for linkage group XIII of the house mouse. *Heredity*, 12, 77-95.
- WALLACE, (≡≡ WRIGHT) M. E. 1947. Two sex linkages in the house mouse with unusual recombination values. *Heredity*, 1, 349-354.
- WALLACE, M. E. 1957. A balanced three-point experiment for linkage group V of the house mouse. *Heredity*, 11, 223-258.
- YATES, F., AND COCHRAN, W. G. 1938. The analysis of groups of experiments. *J. Agric. Sci.*, 28, 556-580.