A Quantitative Method for Analysing Competition where Individuals Cannot be Identified

David A. Hay
Department of Genetics and Human Variation, La Trobe University, Bundoora, Vic. 3083.

Abstract
Methods of analysing the genetic diallel cross, for which different genotypes are mated in all possible combinations, have been modified for the study of competition or other interactions between individuals or between subgroups, even if it is impossible to identify each participant. Two alternative analyses of variance provide a test for any overall effect on performance through combining in a group individuals differing in some respect, as well as distinguishing between more complex forms of interaction. In addition, analysis of the variances and covariances among the different groups can show (1) if there is any consistent interaction, (2) if any divergence from such consistency is due to a single type of individual and (3) if any correlation exists between the magnitude of the interaction and the performance of the individuals tested separately, either on the trait under consideration or on any other behavioural or physical variable.

The method of analysis is illustrated by data on the interactions between females of six inbred strains of *Drosophila melanogaster*, using as measures preening behaviour and the distances between individuals. Suggestions are made about the application of the method to other aspects of social behaviour.

Introduction

Competition within and between plant species is normally studied by growing the plants in all possible pairwise combinations. The interaction between the members of each species can be likened to the hybrids from crosses between pairs of individuals and can be analysed in a similar manner to the genetic situation. Thus Durrant (1965) and Norrington-Davies (1967) applied two techniques of biometrical genetics, the analysis of variance devised by Hayman (1954) and the examination of variances and covariances (Jinks 1954), to the differences between pairs of species in competition. These two biometrical techniques are designed for the diallel cross, where all possible matings are made between a variety of different genotypes in order to determine the genetic control acting within the entire population. Several others, such as Eberhart *et al.* (1964) and Williams (1962), have derived analyses of the competition diallel cross that have many similarities with the Hayman analysis except that their reasoning and nomenclature were developed specifically for a competition rather than a genetic context.

Retaining the original genetic notation wherever possible, the present paper explains the use of the Hayman and other analyses of variance as well as the variance-covariance analysis in providing a more comprehensive description of the competition and other forms of interaction that occur in experimental situations where the individuals cannot be distinguished. In plant competition experiments there is rarely any problem in distinguishing the species, even when they are grown in the same plot. In this case
the yield, as measured from the half-plot, is used in calculating the effects of competition and the existing modifications by Durrant (1965) of the Hayman and variance–covariance analyses rely mainly upon these half-plot effects. But in the genetic situation it is usually impossible from a single hybrid to recognize such differences and separate the contributions of the two parents. Likewise the dyad here has been treated as a single unit without any attempt to distinguish between the individuals in the encounter. Parson (1965) also used a similar but simpler form of genetic analysis to assess the relative importance of the male and female partners in determining mating speed in Drosophila melanogaster.

In general, not having to recognize specific individuals makes most methods of genetic analysis very useful in studying the behaviour of insects, as well as several of the higher animals. Although the details of each participant in an encounter may be known to the observer, there may be no features by which they may be distinguished individually once in the group, without artificial marking which may itself influence the results.

Analysis of Variance

For convenience we shall refer throughout this paper to interactions between individuals of different ‘types’, whether they be plants of different species or, as in the example given to illustrate the analysis, fruit flies of several genotypes, known to differ on observable behavioural indices.

Suppose that one can devise a situation where the performance of all possible pairwise combinations of \( n \) different types can be studied, so that the data can be written as the \( n \times n \) triangular matrix:

\[
\begin{array}{cccc}
\text{Type} & 1 & 2 & \cdots & n \\
1 & Y_{11} & Y_{12} & \cdots & Y_{1n} \\
2 & Y_{22} & \cdots & & Y_{2n} \\
\vdots & & & & \vdots \\
n & & & & Y_{nn}
\end{array}
\]

where \( Y_{rs} \) is the score of the dyad involving the types of the \( r \)th row and \( s \)th column. Each type must be tested with every other type, including once with another of the same type (the diagonal entries). In certain situations, it may be more realistic to enter on the diagonal the score of each individual tested in isolation, since this provides an essential control in determining the influence of group factors.

Assuming that we can test each type with another of the same type (the precise model, but not the method of analysis, differs slightly if the diagonal elements represent instead the scores of isolated individuals), then the score \( Y_{rs} \) can be partitioned into

\[
Y_{rs} = m + X_r + X_s + X_{rs} + \epsilon_{rs},
\]

where \( m \) is the overall mean and \( X_r \) and \( X_s \) are half the deviations from \( m \) of the \( r \)th and \( s \)th types when tested separately, so that

\[
Y_{rr} = m + 2X_r + X_{rr} \quad \text{with} \quad \sum_{i=1}^{n} X_i = 0.
\]
It makes no difference here whether the types are regarded as fixed or whether they are a random sample from those available in a population.

Apart from \( e_{rs} \) (the residual error), the remaining variation is the item \( X_{rs} \), which represents the interaction between the two types and can take one or more of the following three forms:

1. A simple increase or decrease in the means of all combinations compared with the average of the types tested separately. Eberhart et al. (1964) refer to this as \( k \), the general competition effect, and it is mentioned but not given a specific name by Williams (1962, pages 518, 520).

2. Effects unique to the \( r \)th and \( s \)th types and resulting from the introduction of any other type. These effects are additional to (1), the directional effect common to all types, and sum to zero. Williams calls this \( H \), but Eberhart's term \( sk \) is more obvious—it is the difference between types \( (s) \) in their value of \( k \), the common directional effect.

3. Any interaction between \( r \) and \( s \) specific to this particular combination of types, called \( c \) by Eberhart.

To analyse these interactions we can consider firstly two alternative ways of partitioning the total sum of squares between the elements of the \( n \times n \) triangular matrix defined above. This equals

\[
\sum_{r=1}^{n} \sum_{s=r}^{n} Y_{rs}^2 - \frac{2}{n(n+1)} \left( \sum_{r=1}^{n} \sum_{s=r}^{n} Y_{rs} \right)^2,
\]

with \( \frac{1}{2}(n+2)(n-1) \) degrees of freedom. Though the items within each of these partitions given below are orthogonal, certain comparisons between mean squares obtained from different analyses provide variance ratio tests for the interactions. However, these tests often involve a very limited number of degrees of freedom and, in the case of the fixed-effects model, it is advisable to have independent replicates, so that each item may also be tested against its replicate interaction. Such replicate error terms have an additional benefit of relevance to the larger error variation commonly found when types are combined, because of the many more variables involved than when they are tested separately. In his analysis of data on plants, Durrant (1965) showed how a spurious indication of competitive effects can result from this difference in variability. The use of a test of significance based on its own replicate interaction effectively weights each item by its error.

The first partitioning that may be made is into the following three components of variation:

1. The variation between the types tested separately, i.e. the values along the leading diagonal. This item, the 'diagonal' sum of squares, is

\[
\sum_{r=1}^{n} Y_{rr}^2 - \frac{1}{n} \left( \sum_{r=1}^{n} Y_{rr} \right)^2,
\]

with \( n-1 \) degrees of freedom, and it tests the variation between the \( n \) types, ignoring their performance in combination with the others.

2. The variation between all the combinations. Any interaction will affect the magnitude of this 'off-diagonal' sum of squares of

\[
\sum_{r=1}^{n} \sum_{s<r}^{n} Y_{rs}^2 - \frac{2}{n(n-1)} \left( \sum_{r=1}^{n} \sum_{s<r}^{n} Y_{rs} \right)^2,
\]

which has \( \frac{1}{2}(n+1)(n-2) \) degrees of freedom. In the absence of any interaction, the
range of these values must be less than that of the diagonal elements, because all
combinations will lie between the two most extreme scores of the types tested separ-
ately. If the off-diagonal, but not the diagonal, mean square is significant, then there
must be interaction increasing the variation between combinations, for which a more
stringent test is provided by the variance ratio of the off-diagonal to the diagonal
mean squares. Since the numerator and the denominator can be reversed to study the
situation where interaction reduces the variability, this test is two-tailed, unlike the
normal variance ratio test.

3. The third item, with one degree of freedom, tests the difference between the
overall diagonal and off-diagonal means. It may be obtained either by subtraction of
(1) and (2) from the total sum of squares or by the formula for the $b_1$ mean square in
the alternative partitioning considered below.

The above analysis of variance was proposed and used by Jinks and Mather (1955)
for the genetic diallel cross, but it can only reveal the presence of interaction between
the types and is not a test for whether they differ in their interaction. Instead, we may
reinterpret the terms of an alternative analysis of variance developed by Hayman (1954)
and subsequently modified by Morley Jones (1965) for the half-diallel table, which
uses only the triangular rather than the full square matrix, thus disregarding any
differences between using one type as male rather than female parent. A fuller account
of the use, particularly in behaviour genetics, of the Hayman analysis and the variance–
covariance analysis, to be considered later, can be found in Broadhurst (1960).

The components of variation in the Hayman analysis are:

1. The 'array' sum of squares. The diagonal mean square estimates the variability
between $n$ types tested individually, but in addition one can take into account all the
combinations where each type occurs, so that, rather than comparing $Y_{rr}$ values, one
compares the arrays

$$
\left( \sum_{s=r}^{n} Y_{rs} + \sum_{s=1}^{r} Y_{sr} \right),
$$

where the values for type 1, type 2, · · · are obtained from the matrix as

$$
\begin{bmatrix}
2 & 1 & 1 & \cdots & 1 \\
0 & 0 & \cdots & 0 \\
\vdots & & & & \\
0 & & & & \cdot \\
\end{bmatrix},
\begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
2 & 1 & \cdots & 1 \\
\vdots & & & & \\
0 & & & & \cdot \\
\end{bmatrix},
$$

In Morley Jones’ terminology, the array sum of squares is calculated as

$$(dev)^2 u_r/(n+2),$$

where $(dev)^2$ is the sum of squares of deviations about their mean, and

$$u_r = \sum_{s=r}^{n} Y_{rs} + \sum_{s=1}^{r} Y_{sr},$$

with $n-1$ degrees of freedom. Where all mixed and unmixed groups comprise equal
numbers of individuals, i.e. where the diagonal elements are not the score of the
individual by itself but when with others of the same type, we are considering in the
diagonal mean square the variability between the values $2X_r$ ($r = 1 \cdots n$), and in the
array mean square that between the values $2X_r+(n-2)(X_r+sk_r)$, where $sk_r$ is the second form of interaction, namely the effect, assessed over all types, of testing the $r$th type in combination with the others. Therefore, a straightforward variance ratio comparison between the array and diagonal mean squares will indicate the presence of interaction. This test is also two-tailed, because interaction may either increase or decrease the variability between the types, and, since each mean square has only $n-1$ degrees of freedom, interaction will have to be very extensive before this ratio is significant, unless many types are tested.

The remaining sum of squares in the Hayman analysis comprises the $b$ item, which can be partitioned into three orthogonal tests for the different types of interaction explained earlier.

2. The $b_1$ sum of squares of

$$\frac{1}{n(n^2-1)} \left( 2 \sum_{r=1}^{n} \sum_{s=r}^{n} Y_{rs} - (n+1) \sum_{r=1}^{n} Y_{rr} \right)^2,$$

with one degree of freedom, provides an unambiguous test for any directional interaction common to all combinations.

3. The $b_2$ item assesses the variability between types in the difference between their array and diagonal means, so that the values for type 1, type 2, \cdots are obtained from:

$$\begin{bmatrix}
-n+1 & 2 & 2 & \cdots & 2 \\
-1 & 0 & \cdots & 0 \\
-1 & 0 & \cdots & 0 \\
& \vdots & & \ddots & \\
-1 & & & & -1 \\
\end{bmatrix}
\begin{bmatrix}
-1 & 2 & 0 & \cdots & 0 \\
-n+1 & 2 & \cdots & 2 \\
-1 & \cdots & 0 \\
& & & & \vdots \\
\end{bmatrix} = \begin{bmatrix}
1 & -1 & -1 & \cdots & -1 \\
-1 & 2 & 0 & \cdots & 0 \\
2 & \cdots & 0 & \cdots & 0 \\
& \vdots & \ddots & \cdots & \vdots \\
-1 & & & \cdots & -1 \\
\end{bmatrix},$$

allowing for $b_1$, which is included in these comparisons, the $b_2$ sum of squares can be obtained most conveniently as $(\text{dev})^2 t_s/(n^2-4)$, where $t_s = 2u_r - (n+2)Y_{rr}$, with $n-1$ degrees of freedom, and tests whether the types differ in the magnitude of their interaction. Although this test appears therefore less ambiguous than that of the array mean square for the detection of type differences in interaction, the latter has one advantage. If one considers again the additions made by any interactions to the array, compared with the diagonal values, it is clear that where the interactions $(sk_r)$ are correlated with the scores when tested separately $(X_r)$, the differences between the arrays will be inflated and the array mean square will be more likely to reveal interactions than the $b_2$ mean square.

4. The $b_3$ sum of squares, with $n(n-3)$ degrees of freedom, is obtained by subtraction of items (1), (2) and (3) from the total sum of squares and indicates how much of the interaction is unique to combinations between particular pairs of types.

The Hayman analysis can be seen to provide more information than the first partitioning, but it offers no test of the relationship between the performance of the types tested separately and the extent of any interaction, except by inference from the relative magnitude of the array and $b_2$ mean squares, as mentioned above. This information may be obtained and the interactions specified more precisely by examining the variances and covariance of the arrays.
Variance–Covariance Analyses

The role of the variance–covariance analysis in detecting interaction between genes at different loci is fully described in Mather and Jinks (1971) and only a brief summary of the rationale and calculations will be given here. Within each array, two second-degree statistics can be calculated, namely $V_r$, the variance between the means of all the pairwise combinations making up the array, and $W_r$, the covariance of these means with the diagonal entries, i.e. with the scores in isolation of each type with which $r$ is combined. Thus, for a situation with $n$ types, the values for array 1 would be

$$V_1 = \frac{1}{(n-1)} \left( \sum_{i=1}^{n} Y_{ii}^2 - n^{-1}\left( \sum Y_{ii} \right)^2 \right),$$

and

$$W_1 = \frac{1}{(n-1)} \left( \sum_{i=1}^{n} \left( Y_{ii} \times Y_{ii} \right) - n^{-1}\left( \sum Y_{ii} \times \sum Y_{ii} \right) \right).$$

Notice that the value $Y_{ii}$ is only counted once, while $2Y_{ii}$ was used previously in calculating $u_i$ for the array sum of squares.

The type that is least influenced by the presence of other types will have the smallest value of $V_r$, because its performance changes least over all the combinations. Similarly, it will have the smallest covariance with the performance in isolation of the types with which it is combined, i.e. the smallest $W_r$ value. $V_r$ and $W_r$ are thus correlated statistics and, from a linear regression of $W_r$ on $V_r$ over all $n$ arrays, we can recognize three sorts of interaction:

1. No interaction. All the values ($V_r, W_r$) will lie at the same point and there will be no significant regression. As usual in regression, the slope ($b$) has $n-2$ degrees of freedom.

2. A form of interaction which is consistent for all combinations within an array. In this situation, it can be shown that the value $(W_r - V_r)$ should be consistent over all arrays and there will be a linear regression of $W_r$ on $V_r$ with a slope of $+1.0$.

3. Interactions restricted to certain of the combinations involving a particular type. This sort of interaction will cause the variance of that type to increase while the covariance decreases, with the result that the regression slope falls below $+1.0$. If one type is so divergent from the others in the form of interaction that it can be detected from the graph of $W_r$ against $V_r$, then the regression may be recalculated, eliminating all combinations involving this type. Further details of this elimination procedure are given by Jinks (1954) for the genetic diallel and by Harper (1963) for the competition diallel (in his case, competition amongst varieties of linseed and flax).

Since the type most influenced by the presence of other types has both the largest variance and covariance, regressing the sum $(W_r + V_r)$ for each array onto the corresponding diagonal element of the matrix will indicate whether or not the magnitude of any interaction is related to the performance of the types considered separately. Furthermore, the $(W_r + V_r)$ sums may be regressed onto scores for the types on other traits, to detect any relationship of these traits with the extent of interaction. If the data have been obtained from replicate blocks, then the regressions of $W_r$ on $V_r$ and $(W_r + V_r)$ on the diagonal elements can be supplemented by analyses of variance of $(W_r - V_r)$ and $(W_r + V_r)$, using the replicate interactions as error terms (Allard 1956).
Example: Interaction between Six Strains of *D. melanogaster*

*D. melanogaster* is a convenient organism in which to investigate the genetic control of behaviour, and the present data were obtained during a series of genetic analyses of locomotor activity and preening in this insect. This particular experiment was concerned to discover if flies could distinguish individuals from strains other than their own and is fully described in a complementary paper (Hay 1972a).

Two behavioural traits were studied, which were already known to be influenced by social factors: (1) preening of wings, legs and body surfaces, a behaviour pattern whose probable function is to fend off other flies (Connolly 1968); and (2) the mean distance, calculated over all members of a group, between each fly and the individual nearest to it. This measure was developed initially by Clark and Evans (1954) for use in plant ecology and has been applied to *Drosophila* by Sexton and Stalker (1961), who demonstrated that the distribution of flies ceases to be random only when they are crowded. In planning the present experiment it was found that 10 flies were the maximum number that could conveniently be studied in the 18-mm-diam. dishes that served as observation chambers. Only females were tested, because of the problem of pseudocourtship between grouped males (Connolly 1968).

Four different conditions of rearing were used, but these have been summed together for the purposes of the present analysis. In all, 168 groups of 10 female flies were tested, consisting of eight sets (two replicates from each of the four conditions) of the 21 possible pair combinations of five flies from the six inbred strains, Oregon, Samarkand, Florida, 6C/L, Edinburgh and Wellington.

The 10 flies in each group were introduced into an 18-mm-diam. observation dish using light CO₂ sedation, the dishes were assigned randomly to positions on a baseboard and, after 30 min recovery time, the flies were roused by a tap on the base. The number of flies preening in each dish was recorded 15, 30, 45 and 60 s after this stimulation. The data for the analysis consisted of the proportion of the 10 flies which were preening, summed over the four observations and transformed to angles by the formula for \( n = 40 \) of Mosteller and Youtz (1961).

From photographs taken at the beginning, middle and end of this observation period, that is at 0, 30 and 60 s the distribution of flies was scored by measuring the distances between the thoraces of the nearest flies for the 10 individuals and three frames of each group.

**Table 1.** Preening among six strains of *D. melanogaster*

Proportion of flies preening (angular transformation) summed over four conditions of rearing and two replicates. \( u_r, t_r, W_r \) and \( V_r \) are as defined in the text

<table>
<thead>
<tr>
<th></th>
<th>Oregon</th>
<th>Samarkand</th>
<th>Florida</th>
<th>6C/L</th>
<th>Edinburgh</th>
<th>Wellington</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>180·64</td>
<td>166·70</td>
<td>187·09</td>
<td>197·69</td>
<td>202·05</td>
<td>208·59</td>
<td>208·59</td>
<td>Oregon</td>
</tr>
<tr>
<td>145·12</td>
<td>176·38</td>
<td>152·96</td>
<td>192·65</td>
<td>193·62</td>
<td>219·98</td>
<td>220·96</td>
<td>Samarkand</td>
</tr>
<tr>
<td></td>
<td></td>
<td>173·68</td>
<td>206·07</td>
<td>200·82</td>
<td>189·92</td>
<td>189·92</td>
<td>Florida</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6C/L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Edinburgh</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Wellington</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strain</th>
<th>( u_r )</th>
<th>( t_r )</th>
<th>( W_r )</th>
<th>( V_r )</th>
<th>( W_r + V_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oregon</td>
<td>1323·40</td>
<td>1201·68</td>
<td>286·44</td>
<td>237·52</td>
<td>354·26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>293·50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>647·76</td>
</tr>
<tr>
<td>Samarkand</td>
<td>1167·28</td>
<td>1173·60</td>
<td>440·65</td>
<td>379·13</td>
<td>563·66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>450·68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1014·34</td>
</tr>
<tr>
<td>Florida</td>
<td>1265·44</td>
<td>1119·84</td>
<td>184·30</td>
<td>275·01</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>6C/L</td>
<td>1255·84</td>
<td>1122·24</td>
<td>417·37</td>
<td>485·63</td>
<td>484·99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>381·34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>866·33</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>1463·96</td>
<td>1168·08</td>
<td>276·23</td>
<td>123·27</td>
<td>260·84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89·44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>350·28</td>
</tr>
<tr>
<td>Wellington</td>
<td>1374·66</td>
<td>1229·96</td>
<td>362·90</td>
<td>294·27</td>
<td>459·36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>364·43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>823·79</td>
</tr>
</tbody>
</table>

The preening data are given in Table 1, together with the values of \( u_r, t_r, V_r \) and \( W_r \) for each strain. These statistics were used for the analyses of variance shown in Table 2,
where each sum of squares has been divided by a factor of 8 to permit the 'between-replicates' mean square to be used as error. Since practical considerations made it necessary to score the 168 groups of flies in the experiment in a randomized sequence, rather than in a design where conditions and replicates were cross-classified, no separate replicate interactions were available for testing the main effects. However, the replicate variances in the four conditions were homogeneous and hence could be pooled to give a common 'between-replicates-within-conditions' item with 84 degrees of freedom.

Table 2. Analysis of preening in six strains of D. melanogaster, pooled over four rearing conditions and two replicates

<table>
<thead>
<tr>
<th>Item</th>
<th>Values including Florida</th>
<th>Values excluding Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S.S.</td>
<td>D.F.</td>
</tr>
<tr>
<td>Total</td>
<td>1091.37</td>
<td>20</td>
</tr>
<tr>
<td>Partition 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagonal</td>
<td>370.17</td>
<td>5</td>
</tr>
<tr>
<td>Off-diagonal</td>
<td>684.10</td>
<td>14</td>
</tr>
<tr>
<td>Directional (b1)</td>
<td>37.10</td>
<td>1</td>
</tr>
<tr>
<td>Array</td>
<td>833.39</td>
<td>5</td>
</tr>
<tr>
<td>b</td>
<td>257.98</td>
<td>15</td>
</tr>
<tr>
<td>b1</td>
<td>37.10</td>
<td>1</td>
</tr>
<tr>
<td>b2</td>
<td>36.76</td>
<td>5</td>
</tr>
<tr>
<td>b3</td>
<td>184.12</td>
<td>9</td>
</tr>
<tr>
<td>Error</td>
<td>2640.96</td>
<td>84</td>
</tr>
</tbody>
</table>

Regression of \( W_r \) on \( V_r \): (D.F. = 4) \( +0.5320 \pm 0.2882 \) \( (D.F. = 3) +0.8231 \pm 0.1326** \)

Regression of \( (W_r + V_r) \) on the diagonal means: (D.F. = 4) \(-5.6456 \pm 3.0477 \) \( (D.F. = 3) -8.5736 \pm 2.2013* \)

* \( P < 0.05 \). ** \( P < 0.01 \). *** \( P < 0.001 \).

The significant differences in preening between the six strains (the diagonal mean square) are greatly increased by considering the performance in combinations with other strains, the ratio of array: diagonal mean squares being 2.25. Although it is therefore likely that the effects of social interaction are correlated with the scores for the strains tested separately, the array mean square accounts for so much of the total variation that it is impossible to obtain more information on the forms of the interaction between strains from the partitioning of the \( b \) sum of squares.

The fact that most of the remaining variation is contained in \( b_3 \), the item for interactions specific to certain combinations, can best be appreciated from the variance–covariance diagram in Fig. 1. This regression is not linear because one strain, Florida, shows unusual interactions with the other five lines. If we eliminate Florida and consider only the \( 5 \times 5 \) triangular matrix (the analysis shown on the right of Table 2), the \( b_3 \) mean square is reduced, the array mean square accounts for even more of the variation and the regression of \( W_r \) on \( V_r \) becomes linear, with a slope not differing significantly from \( +1.0 \). In addition, there is a significant negative regression of \( (W_r + V_r) \) on the diagonal scores, indicating that the extent of the interaction is correlated with the performance of the strains by themselves, the behaviour of those
strains normally preening less being most influenced by the presence of flies from other strains.

On the other hand, in the analysis of the data on the distance between individuals, the array mean square was significantly less than the diagonal mean square ($F$ ratio = $7.45$, two-tailed $P < 0.05$, for 5 and 5 degrees of freedom). The small array mean square meant that most of the variation on this measure was included in the $b$ item, but when this was partitioned only the $b_2$ mean square was significant ($P < 0.01$). In the absence any of simple directional effect ($b_1$), a significant $b_2$ term implies strain differences in the influence of other flies, with some strains showing an increase in the average distance between flies and others a decrease. It can be seen from Fig. 2 that the regression of $W_r$ on $V_r$ is neither linear nor does it have a slope of $+1.0$, since

![Fig. 1. Covariances ($W_r$) plotted against variances ($V_r$) for interactions between strains in preening.](image1)

![Fig. 2. Covariances ($W_r$) plotted against variances ($V_r$) for interactions between strains in 'distance to nearest neighbour'.](image2)
the six strains differ considerably in their covariances \( W_r = -33\cdot00 \pm 1506\cdot14 \) but have very similar variances, especially if we omit the discrepant value of 1387\cdot12 for Samarkand. This strain is known to show far less locomotor activity than any of the other lines (Hay 1972b), a factor which may well influence its performance on the measure of dispersion and, excluding Samarkand, we have \( V_r = 3634\cdot80 \pm 147\cdot42 \). Both the small array mean square and this analysis of the variances and covariances would indicate that all groups with flies of two different strains exhibit a similar level of performance, i.e. the effect of combining strains is to reduce the normal strain variation in respect of the preferred inter-individual distance and hence to distribute all 10 flies more equally over the experimental area. The regression of \( W_r \) on \( V_r \) no longer has a slope of +1\cdot0 because, as the number of types is increased, interaction of this form will cause the variances of the arrays to become more alike, but the covariances depend more on the differences in performance between the types tested separately and hence will not tend to similarity so rapidly.

**Discussion**

A detailed discussion and interpretation of these results is provided in Hay (1972a), together with further examples of this type of analysis. The analysis can reveal most forms of interaction between the members of a group and the detection of the reduced variability in the distance measure is particularly noteworthy, because this form of interaction might well otherwise be overlooked in alternative types of analysis. The analogy with the genetic analysis has many merits. One advantage of the diallel cross over other experimental mating programs is that it samples a wide variety of parental types, thereby reducing the reliance on information from any single hybrid, which may be atypical of the form of genetic control commonly found within the population (Broadhurst 1960). Similarly, by detecting any divergent type, the variance–covariance analysis provides a useful safeguard when drawing general conclusions about interactions, even though some other, more detailed, analysis would be needed to determine the nature of this type’s unrepresentative performance, e.g. the statistics Williams (1962) calculated for each of his species combinations, to let him eliminate unique interactions, leaving in the analysis only those species which exhibited consistent competition effects.

Williams’ analysis was developed for plant competition, where individuals can be distinguished and where the basic unit is the ‘half-pot’, i.e. the score of each of the two species grown together in a pot. He considers a ‘main’ competition effect, which occurs equally in within- as well as between-species combinations, while the model presented in this paper assumes only interaction between species, or strains in the case of the example given above. However, it is known that there are interactions between flies of the same strain, in that the behaviour of *D. melanogaster* is affected by an increase in group size (Hay 1973). Some additional data, obtained as part of the experiment from which our example was taken, illustrate how the present analysis can be extended to study such effects within each type. When the six pairwise combinations within each strain of five flies from three replicate cultures were scored, the array mean square for preening was significantly smaller than its strain interaction. This was not simply the result of strain differences in the variability between replicate cultures, for the diagonal mean square was no different from its strain interaction. Rather the reason was that the six strains differ in their behaviour towards flies from other cultures of their own
strain, as a separate analysis of each strain subsequently confirmed. As Allard (1956) demonstrated for the genetic diallel cross, the ability to apply a factorial design in this way to the Hayman analysis of variance, and also to the analyses of variance of \((W_r - V_r)\) and \((W_r + V_r)\), offers tests of significance for changes in the various forms of interaction and competition which are applicable to many experimental treatments, such as the different rearing conditions of our fruit flies.

Federer (1967) describes many other situations where the diallel design and analyses are applicable in psychological and biological testing, e.g. job status evaluation and successive dosing with the same or different drugs. There are many measures where it would be interesting and appropriate to employ this analysis, but rarely are there adequate data. We can separate suitable measures into two types. Firstly, there are those involving 'co-active' social facilitation, a term used by Zajonc (1965) for the effects on performance in situations where individuals are independently engaged in the same activity at the same time and in full view of each other. Thus co-action applies only to behaviours which can also occur in isolated individuals, e.g. eating, where grouping increases feeding in rats, chickens and dogs. With a group of subjects paired in all possible combinations, the diallel analysis could be used to study this sort of interaction in greater detail. Zajonc refers to the work of Chen (1937) on the social enhancement of digging in ants, where the time before digging was reduced and the total amount dug was greatly increased when two ants were working together, compared with the performance of two ants each in isolation. Zajonc used these data to reinforce his thesis that grouping individuals leads to an improvement in performance on familiar tasks, but an impairment in the learning of a new skill (both of these would be indicated in the present analysis by a significant \(b_i\) effect). Sudd (1972), testing a different species of ant in a somewhat different experimental situation, failed to observe the facilitation, at least partly because of the time the ants spent in grooming, food exchange and other behaviours possible only in the grouped situation. Sudd found little relationship between performance in solo and in paired conditions, but he did not try testing the same ants in more than one pair, which would have allowed the diallel analysis with its more detailed breakdown of the interactions between individuals. Such an experimental design would be needed to test Zajonc’s hypothesis, because the facilitation in ants is not constant; Chen (1937) found a differential social effect on the delay before digging, such that grouping had a greater effect on slow than on fast worker ants. One of Sudd’s criticisms is that Chen used no statistical test to check this difference; unfortunately Chen’s published data are inadequate for a diallel analysis, where the \(W_r, V_r\) analysis could provide an unambiguous test for differential interaction. That is, assuming the regression of \(W_r\) on \(V_r\) was linear (showing that the interaction was consistent and not just confined to a few combinations of ants), then the slow workers, being more influenced by their partner, would have high \(W_r\) and \(V_r\) values, which would result in a positive regression of the \((W_r + V_r)\) sums on the length of delay before digging when isolated.

Digging is an example of a behaviour which occurs both in isolated and in grouped ants, but there is a second category of behaviour, including such things as the social grooming of the ants, which can only occur in groups. The chief problem in applying the diallel analysis to such behaviours concerns the diagonal entries in the diallel table. These cannot be obtained from the scores of individuals when alone and are essential for the Hayman analysis. There are alternative statistical analyses for the diallel
cross which do not involve the diagonal entries (Griffing 1956) but, while such analyses could undoubtedly have applications in competition and in behavioural studies, such as in analysing a social dominance hierarchy, they do not provide such a comprehensive partitioning of the variation as the Hayman analysis, e.g. there is nothing corresponding to the $b_2$ term. But if the behaviour cannot be measured in isolated individuals, it may be possible to look at the behaviour when with others of the same group and use these data as the diagonal terms in an analysis of the interaction between groups.

An example of social behaviour relevant to the diallel analysis is aggression between rodent genera, a behaviour which develops before sexual maturity, while intrageneric aggression only develops after maturity (Baenninger 1973). Grant (1970) studied five groups (one Peromyscus species plus mainland and island Microtus and Clethrionomys species) in a large open-field arena, split into grass and woodland habitats. Here he measured habitat preference—the amount of time spent in the grass—and mean number of aggressive encounters. He found that species had a mutually dispersive effect, so that aggressive conflicts were more frequent within than between species. It would be practical to analyse these species interactions more closely by the diallel techniques but, of the two dialles which can be reconstructed from his tables, unfortunately the 5 × 5 diallel for habitat preferences has several species combinations lacking, and the 3 × 3 diallel (one species of each genus) for the number of aggressive encounters is really too small, e.g. there are insufficient degrees of freedom to allow the $b_2$ sum of squares to be calculated, and this would be needed as an error term, in the absence of any other information on the variation within each combination.

Further possibilities for the diallel analysis come from chickens, where aggression can be studied both within and between breeds. Ortman and Craig (1968) examined fights only within and between two selection lines, but their methods could easily be extended to fights within and between a greater variety of groups. A final example concerns aggression in rhesus monkey groups where Marsden (1973) observed that contact between groups decreased aggression within subordinate groups, but increased it in dominant groups. Marsden states (p. 248) that 'there might have been a relationship between specific intergroup behaviours and concurrent intragroup responses'. Clearly this represents a complex experimental situation, but one to which the diallel analysis would seem relevant, in view of the comments made earlier in the Discussion about the use of the diallel to study interaction within, as well as between, D. melanogaster strains.

Finally, it should be stressed that far more comprehensive analyses are possible where individuals can be distinguished, especially in plants which are static and where the spatial relationships between neighbours can therefore be considered. Huh (1972) describes some very detailed analyses of such effects and McGilchrist (1965) has extended the analysis of Williams (1962) to distinguish two forms of competition in plants, the increase in species A when with species B and the associated depression of species B through having A as its neighbour. Perhaps the most extensive analysis developed for plant competition, and one which has many affinities with the present model, is that developed by Durrant (1965) and Norrington-Davies (1967). On the basis of the differences between plant species in their performance in mixtures, they separate competition effects into 'alpha' (where the increase or decrease in performance for any species due to the interaction is constant over all the other species with which it is combined and thus corresponds to the $b_2$ effect here) and 'beta' (the possibly
more realistic case, where the magnitude of any interaction is a function of the difference between that pair of species grown separately. Attempts are being made to extend the current model to the distinction between these two forms of effect, although this has proved difficult where the magnitude of the interaction is correlated with the performance of the types when tested separately. Furthermore, as a final cautionary note, the complex way in which the total variation is partitioned, even with the model in its present state, means that the need both for large samples and reliable measures cannot be overemphasized.

Acknowledgments

This work was performed during the tenure of a Science Research Council postgraduate studentship and forms part of a program of research in psychogenetics in the Departments of Genetics and Psychology at Birmingham University, supported by a grant from the Medical Research Council. The author wishes to thank Professor P. L. Broadhurst for his helpful comments on the manuscript.

References


Marsden, H. M. (1973). Aggression within social groups of rhesus monkeys (Macaca mulatta): 
effect of contact between groups. Anim. Behav. 21, 247–9.
Norrington-Davies, J. (1967). Application of diallel analysis to experiments in plant competition. 
Euphytica 16, 391–406.
Parsons, P. A. (1965). The determination of mating speed in D. melanogaster for various combinations 
of inbred lines. Experientia (Basel) 21, 478.
Sudd, J. H. (1972). The absence of social enhancement of digging in pairs of ants (Formica lemani 
9, 77–81.

Manuscript received 18 September 1973