quotient test ($P = 0.01$); and so on. For the most part, these differences are in keeping with those found in America. Although there are many individual exceptions, they do suggest that young physical scientists tend to be less intellectually flexible than young arts specialists, and more restricted emotionally. On the other hand, there is at this stage no means of telling which boys will eventually do original work—and therefore no empirical proof that original work is more likely to follow one emotional configuration than the other.

L. Hudson

Psychological Laboratory, Cambridge.

[1] Getzels, J. W., and Jackson, P. W., *Creativity and Intelligence* (Wiley, New York, 1962).
[2] Hudson, L., *Nature*, **186**, 413 (1960).

Theoretical and Empirical Confusion Functions

The following table should replace Table 1 of my communication that appeared under the above title on p. 836 of the August 25 issue of *Nature*.

Table 1

	x_1/x_2							
Author	0	0·5	0·67	0·75	0·80	0·83	0·86	0·88
Crossman	0·00	1·00	1·71	2·41	3·13	3·82	4·51	5·18
Welford	0·33	1·00	1·67	2·33	3·00	3·67	4·33	5·00
Hammerton $k = 1$ $k = 1/14$	0·34 1·00	1·00 1·00	1·64 1·00	2·28 1·05	2·92 1·11	3·57 1·25	4·35 1·39	4·86 1·54

I. M. Hughes

War Office Clothing and Equipment
Physiological Research Establishment,
Farnborough, Hants.

STATISTICS

Partition Correlation Matrices for Heterogeneous Quantitative Data

A form of data which occurs occasionally in sociology, and very commonly in ecology, is that in which the attributes, though measurable when they occur, do not run through the whole population, so that the data matrix contains many zeros. Such a population consists in reality of a number of more or less discrete sub-populations, each defined by some only of the attributes, and the application of traditional multivariate methods to such data encounters two main difficulties. First, even though the non-zero values of an attribute are normally distributed, the addition of zeros causes the mean/variance relationship of the whole attribute to approximate to that of a qualitative (0,1) distribution; if component analysis is used in an attempt to separate the sub-populations the normal $(x - \bar{x})/\sigma$ transformation then results in excessive weight being given to the absence of a common attribute or the presence of a rare one. Secondly, if factor analysis with communalities is attempted, factors after the first—since all factors unrealistically involve all attributes—cannot be interpreted. Data which are heterogeneous in this sense are better subdivided on a presence-or-absence basis; and a series of papers from this laboratory[1-6] has explored one statistical technique for this purpose. However, if the data are not intrinsically entirely qualitative this involves discarding information, and there appears to exist no method of assessing the relative importance of the qualitative and quantitative elements of a given set of data. This communication outlines such a method.

Let the set of values for an attribute x, some of which are zeros, be denoted by a vector $[0, x]$, and let m_x be the mean of the non-zero values. Let this vector be regarded as the sum of a qualitative (L) vector $[0, m_x]$, in which all non-zero values are replaced by m_x, and a quantitative (N) vector $[0, (x - m_x)]$. The sum of the squares of the deviations of the N-vector elements from their (zero) mean is the same as that of the non-zero values of x from their own mean m_x; and the sum of the variances of the L and N vectors is identically equal to that of the original vector. The division may thus be regarded as a true partition between the qualitative and quantitative elements of the attribute. If two such vectors $[0, x]$ and $[0, y]$ are so partitioned, it is easily shown that the sum of the four covariances L/L, L/N, N/L and N/N is also identically equal to that between the original vectors. These covariances may therefore be used, with the corresponding L and N variances, to generate a set of four correlation coefficients (indeterminate values being treated as zeros), which in turn define three correlation matrices.

The first of these, L/L, is a normal correlation matrix; it contains all the purely qualitative information, and is in fact the matrix which would have been obtained if the original data had been replaced by [0,1]. The second, N/N, is also a normal correlation matrix, summarizing the numerical relationships between pairs of attributes when, and only when, both are present in the same individual (one common individual will suffice since what is under examination is the extent to which either deviates from its non-zero mean). The third ($L/N + N/L$) is asymmetric, with zeros in the principal diagonal, and we suggest the convention of L for rows and N for columns. This matrix summarizes the extent to which changes in the quantity of one variable are accompanied by the appearance or disappearance of others.

Inspection of the three matrices shows very clearly where most of the information resides; as an objective measure of the total information content of each matrix we tentatively propose to use the sum of the squares of all the correlation coefficients, excluding the principal diagonal; but this is not easy to justify, and we may wish to modify the proposal with more experience. If most of the information is in the L/L or L/N matrices, then subdivision of the population is clearly indicated; only when, after successive subdivisions, the bulk of the information appears in the N/N matrix is factor analysis indicated, and certain of success. The method has proved extremely effective when applied to specially constructed artificial populations, and we hope to use it to analyse data which have proved intractable by traditional methods.

One of us (M. B. D.) acknowledges the support of a Department of Scientific and Industrial Research maintenance award.

W. T. Williams
M. B. Dale

Department of Botany,
University of Southampton.

[1] Williams, W. T., and Lance, G. N., *Nature*, **182**, 1755 (1958).
[2] Williams, W. T., and Lambert, J. M., *J. Ecol.*, **47**, 83 (1959).
[3] Williams, W. T., and Lambert, J. M., *J. Ecol.*, **48**, 689 (1960).
[4] Williams, W. T., and Lambert, J. M., *Nature*, **191**, 202 (1961).
[5] Williams, W. T., and Lambert, J. M., *J. Ecol.*, **49**, 717 (1961).
[6] Lambert, J. M., and Williams, W. T., *J. Ecol.* (in the press).