HAVE COMPUTER, WILL CLASSIFY? A STONE AGE PROBLEM.

P. Callow

Department of Archaeology, University of Cambridge, Downing Street, Cambridge CB2 3DZ

Classification is an essential, and time-honoured, tool of the archaeologist, in that it is a conceptually simple means of reducing initially unmanageable quantities of information to more manageable proportions. In everyday life, classification is an integral part of the model-building with which we come to terms with our surroundings, and it is a natural extension of this process to apply the method to archaeological data, whether at the level of the individual artefact or of the culture.

Side by side with the development of increasingly powerful analytical techniques (often only practicable with the aid of a computer), there arose in the 1960s and early 1970s a growing awareness of the complexity of the classification process. This may be seen as very much a part of the 'New Archaeology', and the literature of the period yields ample evidence of the importance attached to the question of typology by both of the principal camps. The clearest statement demonstrating the need for a shift from empiricism in this respect towards a more logical positivist approach is probably that by Hill and Evans (1972), though it would be impossible to trace the path by which such a position - now held almost universally among Anglo-Saxon prehistorians, though perhaps less evidently so in the Francophone world - was reached. Coupled with the analytical techniques referred to, which call for explicit statements as to assumptions, the consequence has been (ideally) a formalisation of the typological process.

That typology per se in fact remains necessary despite these increased analytical capabilities is demonstrable by purely practical arguments. It is difficult to imagine a feasible method of reducing hundreds of stone tool assemblages to manageable proportions without resorting to something comparable to the method of Bordes (1961), given that they cannot usefully be treated as samples of tidy unimodal, or normal, distributions and so boiled down to a set of means and standard deviations. Partition of some sort, and analysis of the resulting frequencies, is the only reasonable solution whatever interpretation is offered of the results. Where the 'new' typologist differs from the old is in the pressures on him to be explicit about his method of work and to offer some justification for his decisions. His variables must be chosen wth a particular set of questions in mind, as must the algorithm (clustering or otherwise) used to reduce the data. The computer is thus used in an exact equivalent of the traditional typologist's manual sorting of artefacts spread out on a large table, though usually by trading off the pilot sample size in return for increased rigour. The typological scheme thus derived from a preliminary study, the computer can then be used to assign additional pieces to their respective classes, or it may be possible to proceed more or less manually from this point by means of a simple key (especially where the attributes are qualitative). Of course the greatest gains are to be found in the comparison of whole assemblages or even cultures in as much as these lend themselves not at all to physical manipulation!

Given these 'advances' it is hardly surprising that in recent years there has been a burgeoning of computer-aided classificatory studies to the point where even Ph.D. students are liable to feel that they have somehow underachieved unless they have produced something innovatory on the taxonomic front. Certainly the most popular algorithms for multivariate work have proved robust enough to yield results of 'archaeological significance' when used in a fairly simple-minded way. Yet they rarely behave quite well enough to satisfy the expectations of the user; there is almost always the odd case whose unlooked-for position on a dendrogram has to be explained away as the result of the limitations of the algorithm. Perhaps more significantly, the results of such studies do not as a rule lead to the widespread adoption of new typologies in dealing with further material, especially where the study is based on qualitative attributes; this is often a consequence, quite simply, of the complexity of the parameters used in the type definitions, but it also reflects a general lack of confidence. This last arises from three distinct problem areas:

(i) difficulties in selecting 'meaningful' input variables - even when the hypothesis to be tested appears to be fairly straightforward

(ii) uncertaity as to the appropriateness of whichever happens to have been chosen of the wide range of clustering and other algorithms currently available, given that all derive ultimately from relatively simple models

(iii) the poor state of development of signifiance tests associated with such forms of analysis.

To sum up, therefore, it is often hard to avoid a residual uncertainty as to the correct interpretation of all but the most obvious results, and the chief advantages of such procedures are often seen to be not so much the creation of more realistic systems of classification (i.e. reflecting more accurately whatever was going on in the past) but rather the benefits of reproducibility of results and the powerful data display opportunities which they offer.

The somewhat pessimistic tone of the above remarks should not, however, be taken as indicating that the author is disillusioned with the computer as an aid to classification. In fact they reflect a concern that the potential of the computer is rarely exploited to the full in such studies, often because the archaeologist is limited by his own computing skills to the use of a somewhat restricted set of routines available in one or more packages at his home installation. Given some combination of personal programming experience, a well-stocked program library or even a tame programmer, he can elaborate his analytical procedures in order both to arrive at a better understanding of the results of the simpler strategies, and to develope more advanced models which may have greater archaeological interest. A classic example is provided by Newcomer and Hodson's (1973) use of constellation analysis to compare the results of employing different classifcatory schemes on burins from Ksar 'Akil, Lebanon.

In the course of a study of lower and middle Palaeolithic handaxes from northwest Europe (Callow 1977), the author found it necessary to resort to a variety of techniques in order to deal with large numbers of artefacts of a kind which have attracted the attention of traditional and computer-aided typologists and metrical analysts (Alimen and Vignal 1952; Bordes 1961; Roe 1968; Barral and Simone 1971; Cahen and Martin 1972; Doran and Hodson 1975, 241-6; Isaac 1977; Monnier and Etienne 1978). The total data consisted of observations on handaxes from 87 assemblages from Britain and north France (apart from five from Belgium and southwest France included for comparative purposes). Up to 34 attributes, mainly quantitative, were employed, though it was not practicable to record the full set for all of the series studied. As a minimum, those attributes proposed by Roe (1968) and Bordes (1961), with a few others, were taken.

The investigation fell into two parts:

(i) a classic inter-assemblage comparison based on the whole list of sites and on certain subsets according to period and area, with clustering and principal components analysis of type frequencies

(ii) a much more elaborate study of the patterning within relatively synchronous groups.

The initial pilot work was briefly described in Callow (1976), and nothing more will be said here of the actual results of the first half of the project. The second part, which is of greater methodological interest, is referred to later.

The purpose of the remainder of this paper is to present some of the techniques which were employed to gain insights into the behaviour of the data and of the primary analytical methods, not as an exhaustive catalogue of all the approaches possible in such case but in order to demonstrate a few of the possibilities. It should be borne in mind that only quantitative attributes have been used, but equivalent techniques may be envisaged for qualitative and mixed data.

Empirical investigation of a partition

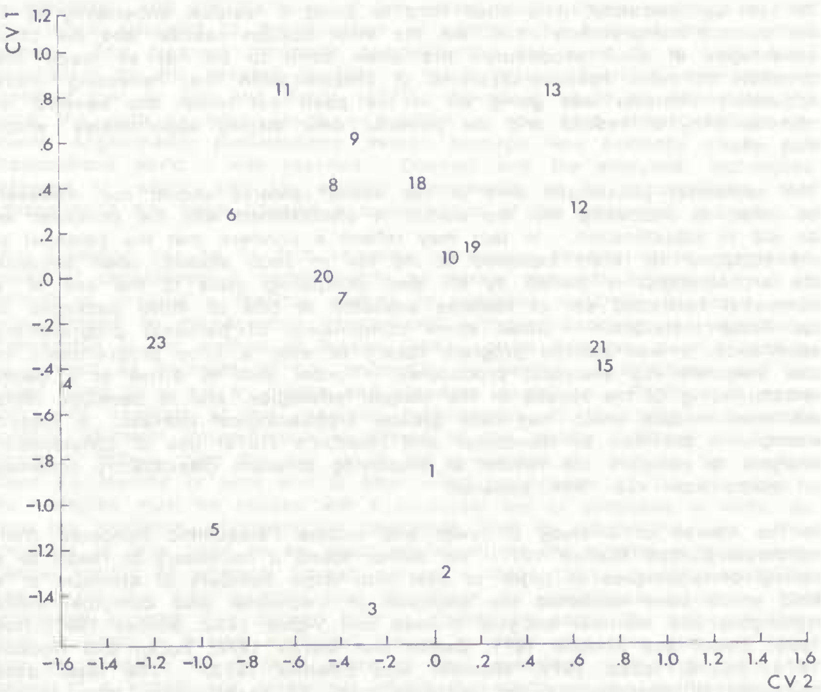

Figure 1: Canonical variate scores (means) for Bordes' handaxe types (identified by numbers).

One of the more interesting aspects of the investigation was that there was already a handaxe typology (that of Bordes) in fairly widespread use. This was intuitively derived, but eventually defined in part by metrical attributes which were extensively employed in partitioning 'thin' handaxes, while 'thick' pieces were classified according to more subjective criteria which in practice are not always easy to apply rigorously. In order to gain an insight into the relationship between the Bordes' typology based on mixed data and a rather impecise algorithm on the one hand, and the purely metrical observations taken for the same pieces, a canonical variates analysis on 5522 pieces was performed, with groups defined by Bordes' types. The resulting type mean scores are plotted in fig 1. It shows that the morphological space into which the artefacts may be fitted according to the metrological criteria has been more heavily partitioned in some areas than in others. This remains true even if the least metrically defined types (13, 18, 19, 20, 21), which include Abbevillian and partly bifacial pieces, are removed from the diagram, and reflects the detailed subdivision of the ovate to cordiform range contrasted with the very much looser classification of the more pointed pieces (1 – 5). The presence of such a bias, irrespective of whether it is justified or not, has to be taken into account when using this typology. In practice, it obscures important variability among the pointed, heavy-butted handaxes when applied to some of the specialised British industries (which, to be fair, had not been seen by Bordes when he set up the scheme).

Elaborated graphical output

A frequent area of uncertainty is the 'reasonableness' of the variable list selected for consideration. This is entirely subjective, of course – that is the root of the whole problem. Nevertheless, most researchers would want an assurance that nothing was going on that was patently ridiculous. A very satisfactory means of achieving this is by labelling a dendrogram, or the points on a scattergram of MDSCAL, principal components or co-ordinates scores, with scaled-down sketches of the items being classified. If a digitiser is available this may be done automatically. Alternatively, photo-reduced drawings may be glued on manually. The drawings may be either naturalistic or symbolic representations (the latter are obviously preferable where qualitative attributes are being used), or even a mixture. Whatever course is followed, the sketches should convey as much as possible of the input information.

Where the objective is to be reached by a reduction of dimensions, and the 2D results are hard to read, or involve a lot of 'strain', it is always worth considering the use of stereo pairs of scattergrams to give a 3D representation.

Stability of Principal Components

The main inter-assemblage investigations employed a modification of Bordes' typology, using 16 entirely metrically defined types, plus a subjective 'cleaver' category. His original measures and threshold values were extended to cover the whole morphological range so that classification could be automated (it should be stressed that, as with the Bordes' system itself, the resulting types were not claimed to have any significance in terms of function, 'mental templates', etc.). The assemblage type-frequencies were then investigated using a number of techniques.

It was rapidly appreciated that, as with the Bordes' typology, the most critical partitioning variable was the ratio breadth/thickness, for which a threshold value of 235 was the 'official' arbiter between 'thick' and 'thin' handaxes. Since the threshold was near the mean of a distribution which was not demonstrably multimodal, it seemed likely that the precise value chosen could be of importance. The analyses were repeated, therefore, with the critical B/T ratio set first at 2.10, then at 2.60. Some changes were apparent in the dendrograms etc., but the most marked differences were in the coefficients defining the principal components. These differences may be expressed in terms of the angular displacement of the principal axis while the basis for partition of the handaxes (and hence for calculating type-frequencies) was undergoing modification (fig. 2). The displacement varied from a minimum of 21 degrees in the case of a subset consisting of all of the French assemblages (both Acheulian and Mousterian) up to 88 degrees for the Mousterian series on their own. The latter in fact implied a complete reversal of the order of the principal components.

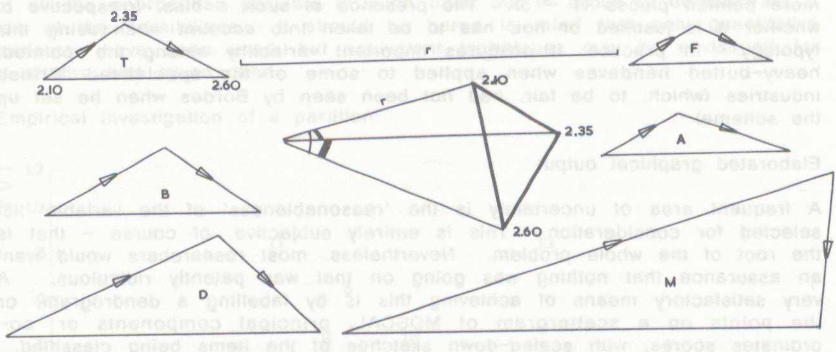

Figure 2: Angular displacement of first principal component, and of the Acheulian-Mousterian discriminant function, as the critical B/T ratio varies between 2.10 and 2.60. The sides of the triangles record chord length (compare with unit radius). The larger the triangle the less stable are the principal axes. Key: T = total sample; F = French sites; B = British sites; A = Acheulian; M = Mousterian; D = discriminant function.

The main conclusion drawn from this study was that Bordes' original choice of a ratio of 2.35 behaved fairly well when the objective was to contrast Acheulian and Mousterian sites as well as maximising the variability within the Acheulian. For investigating variability within the Mousterian, on the other hand, it had the grave disadvantage of forcing into separate types the thicker pieces which were all too evidently better seen as forming the tails of a number of unimodal distributions of predominantly thin pieces. In a separate study, referred to below, it was in fact concluded that handaxe profile data were best ignored in the Mousterian. Or to put it another way, no typology is likely to be very effective across this time-range.

Development of an alternative model

Some of the disadvantages of a traditional typological approach were referred to in an earlier paper (Callow 1976). One of the most severe is the imposition of a static model (static in that the partitioning remains fixed) on the results of a dynamic process. In some respects to limit oneself to a methodology which has its roots in the precomputer era of archaeology is to fail to respond to the opportunity to explore new theoretical possibilities.

In the case of the handaxe data under discussion, it was recognised that no existing typology had attempted to come to grips with the intentions of the makers. Reliable functional data was effectively lacking, the first serious studies were under way at the time (Keeley 1980), but it seems most likely that, apart from some obviously specialised forms, handaxes must have served as multi-purpose tools. A further consideration was the span of time during which handaxes were manufactured; even if they fall within certain limits of morphology, is it reasonable to treat them as if their makers' goals remained unchanged? Also, just as important, what is one measuring when comparing assemblages on the basis of type frequencies (or presence-absence)? Is not such an approach liable to confuse effects which are 'cultural' in origin with those which are the result of site use?

An alternative model was therefore considered, in which the most important matter for consideration was no longer the relative abundance of different types. Instead the handaxes from each series were subjected to a variety of taxonomic experiments to arrive at the best partition of each series, the number of taxa being determined by inspection of the dendrograms, multivariate F-ratios and Marriott's (1971) covariance determinant ratio (fig. 3). This gave, as rule, three or four 'types' for each series. The 'type' means for several asemblages (chosen to cover a limited period) were then plotted on to a single principal components or canonical variates scattergram.

If we suppose that there is no clear internal structuring within each assemblage, the result of such a process, when applied to several assemblages, should be equally unstructured. If on the other hand the 'types' are not the result of arbitrary zonation of a continuum, the consequence should lie between two extremes. Firstly, a similar lack of patterning would indicate that the 'mental templates' being used in tool manufacture, though clear enough in themselves, were not the same from series to series. Secondly, if the 'type means' themselves clustered, it might indicate that the the set of 'mental templates', while not necessarily identical for every series, was nevertheless very similar (fig. 4). This provides a means of assessing cultural affinity on what are essentially stylistic grounds. Moreover, if a site fails to provide a member for one of the clusters on such a diagram it does not disqualify that site from membersip of a postulated group of sites with shared typological aims.

The results of two such experiments are given here, since they provide an interesting contrast which perhaps throws light on the intellectual development of early man.

Fig. 5 shows a scattergram derived from Mousterian handaxes dating to the first half of the last glaciation. Initial experiments were surprisingly unsuccessful, given that by this period one would expect a fair degree of standardisation in toolmaking. Bearing in mind the essentially 'thin' nature of Mousterian handaxes, therefore, all varables relating to profile were omitted and the programs run again. This time the 'type' means exhibit marked clustering, suggesting that previously the structure had been obliterated by noise from the irrelevent profile data. It seems that Neanderthal man made his handaxes with two-dimensional forms in mind corresponding to a number of simple geometrical shapes.

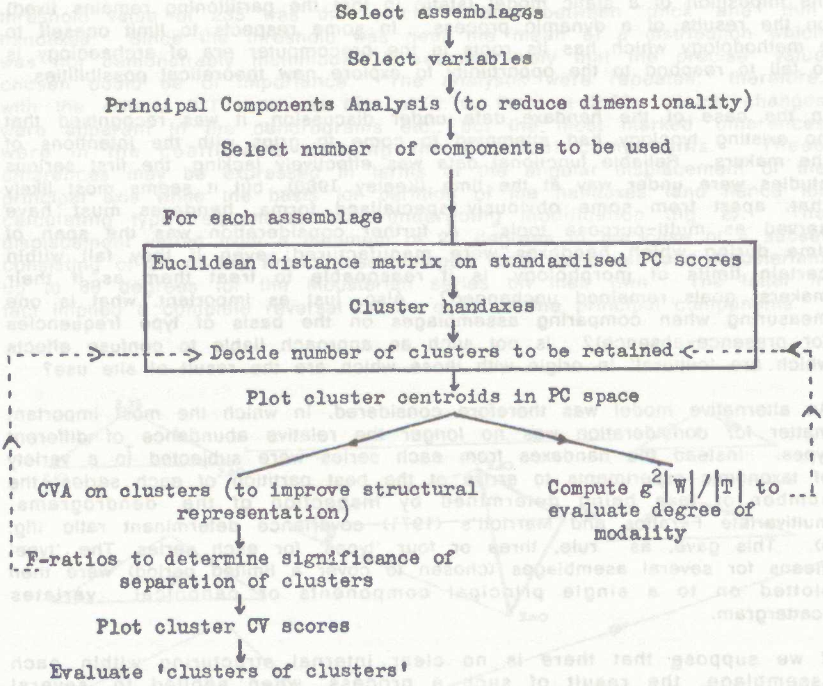

Figure 3: Computational operations used in searching for repetition of 'natural' structuring in handaxe series.

Fig. 6 includes only relatively early (putatively pre-Hoxnian) assemblages. Again, first attempts were disappointing, and the study was almost abandoned on the grounds that perhaps the knappers' goals were too imprecise at this date. In this case, remarkably clear structuring was obtained in due course by omitting data relating to elongation. This somewhat surprising discovery (handaxe length/breadth ratios have figured in the metrical repertoire of all workers in this field) becomes more understandable when the variables most highly correlated with the canonical variates are considered in detail. They relate essentially to pointedness (which probably has the greatest significance in determining which part of the tool provides the working edge), thickness and butt shape (presumably affecting the manner in which the piece is held). These are evidently of functional importance. On the other hand, elongation is liable to be reduced by resharpening and is less obviously functional (except in extreme forms such as 'ficrons' which are lacking at this date). The knappers' goals in this case appear to relate not to the details of shape but to the criteria required to make the tool do its job.

Such an approach can certainly be developed further; the handaxe data, coming from isolated and only loosely dated sites, could usefully be replaced by artefacts from a long stratified sequence, while more recent industries should exhibit clearer standardisation.

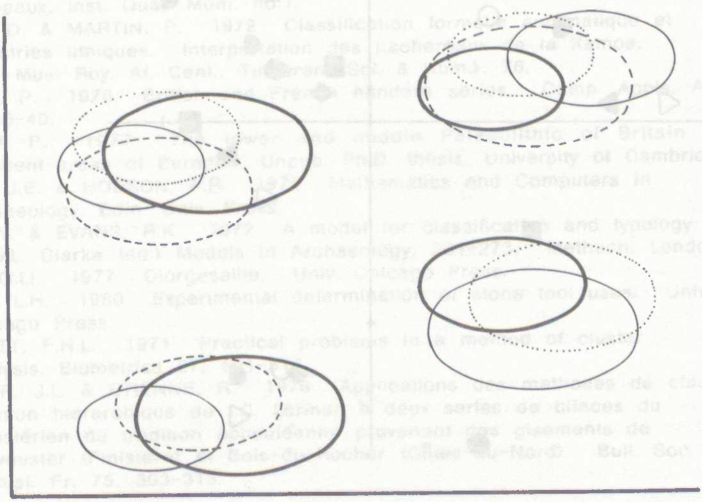

Figure 4: Distribution of artefact groups from four imaginary assemblages (identified by outline type of confidence ellipses) sharing similar knapping goals ('mental templates'). Four clusters are apparent; note that they are not represented in every assemblage.

Conclusion

Each new problem of taxonomy should be treated with caution, and not put through a cookery-book procedure; above all, the difficulties should not be underestimated. It is a consequence of past human ingenuity that no two sets of archaeological data ever behave in quite the same way - some clusters are more elliptical than others, sometmes a single measurement may have quite a different meaning in a new context. This presents the archaeologist with a challenge, given that the theorists have pointed out the dangers of traditional, empirical pathways, while the methodology enabling him to harness the growing power of electronic computers is in its infancy.

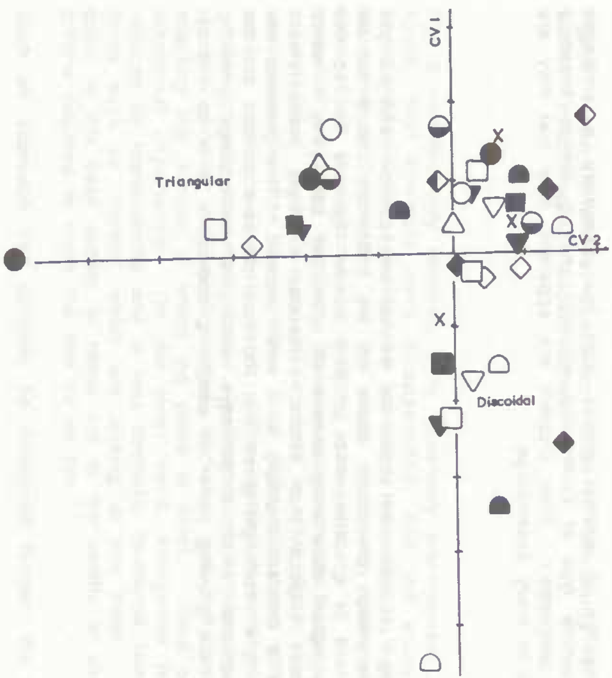

Figure 5: Mousterian handaxe assemblages. Canonical variate scores for centroids of handaxe groups generated by clustering within assemblages (groups from one site share a common symbol).

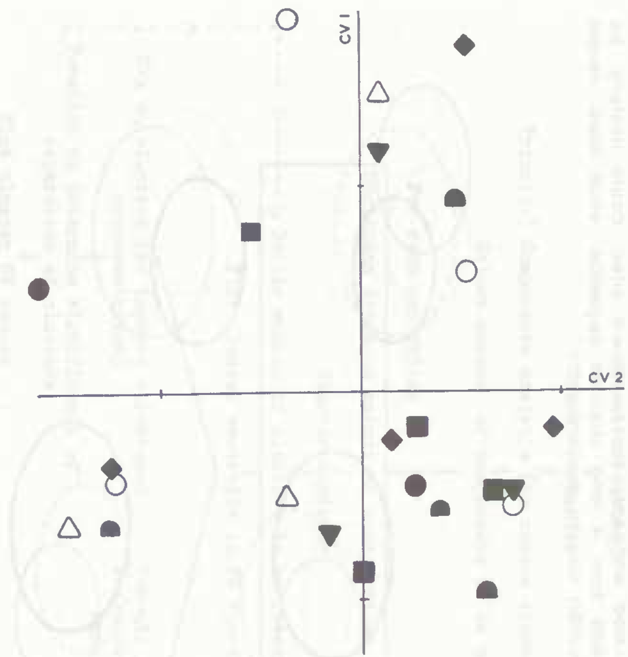

Figure 6: Early Acheulian handaxe assemblages (cf. fig. 5).

ALIMEN, H. & VIGNAL, A. 1952 Etude statistique de bifaces acheuléens - essai d'archéométrie. Bull. Soc. Préhist. Fr. 50, 56-72.

BARRAL, L. & SIMONE, S. 1971 Classification automatique des industries préhistoriques. Les bifaces. Bull. Mus. d'Anth. Préhist. Monaco, 17, 5-38.

BORDES, F. 1961 Typologie du Paléolithique ancien et moyen. Univ. de Bordeaux, Inst. Quat. Mém. no.1.

CAHEN, D. & MARTIN, P. 1972 Classification formelle automatique et industries lithiques. Interprétation des hachereaux de la Kamoa. Ann. Mus. Roy. Af. Cent., Turveren (Sci. & Hum.), 76.

CALLOW, P. 1976 British and French handaxe series. Comp. Appls. Arch. 3, 33-40.

CALLOW, P. 1977 The lower and middle Palaeolithic of Britain and adjacent areas of Europe. Unpub. Ph.D. thesis, University of Cambridge.

DORAN, J.E. & HODSON, F.R. 1975 Mathematics and Computers in Archaeology. Edin. Univ. Press.

HILL, J.N. & EVANS, R.K. 1972 A model for classification and typology. in D.L. Clarke (ed.) Models in Archaeology, 231-273. Methuen, London.

ISAAC, G.Ll. 1977 Olorgesailie. Univ. Chicago Press.

KEELEY, L.H. 1980 Experimental determination of stone tool uses. Univ. Chicago Press.

MARRIOTT, F.H.L. 1971 Practical problems in a method of cluster analysis. Biometrics 27, 501-514.

MONNIER, J.L. & ETIENNE, R. 1978 Applications des methodes de classification hierarchique de I.C. Lerman à deux series de bifaces du Moustérien de tradition acheuléenne provenant des gisements de Kervouster (Finistère) et Bois-du-Rocher (Côtes-du-Nord). Bull. Soc. Préhist. Fr. 75, 303-313.