THE QUANTITATIVE ASSESSMENT OF TAXONOMIC RELATIONSHIPS IN PRIMATE PALAEONTOLOGY*

W. E. LE GROS CLARK, F.R.S.

Department of Anatomy, University of Oxford

ABSTRACT—One of the ultimate aims of the taxonomist, in the assessment of phylogenetic relationships, is to express the latter in quantitative terms. But these aims are likely to be stultified unless due account is taken of the morphological fundamentals of the measurements which are employed for quantitative comparison. This paper is concerned to point out the serious fallacies which are sometimes involved in attempts at quantitative assessment of relationships, particularly in problems of palaeoanthropology.

INTRODUCTION

APART from the problem of assessing general taxonomic relationships by reference to morphological resemblances so far as these may be determined by direct comparisons, attempts have from time to time been made to estimate degrees of resemblance (and thus, it is assumed, degrees of affinity) on a quantitative basis. The biometrical approach is an attempt to facilitate and place on a strictly objective basis the comparison of one type with another. But unfortunately it is fraught with the greatest difficulties, the main one of which, no doubt, is the impossibility by known methods of weighing each individual character according to its taxonomic relevance. If the measurements of every single morphological character of skull, dentition and limb bones were of equal value for the assessment of zoological affinities, it might be practicable to assess the latter in strictly quantitative terms. But it is very well recognized that this is by no means the case. It is well known also, that the products of convergent evolution may lead to similarities (particularly in general over-all measurements and indices derived therefrom) which if expressed quantitatively would give an entirely false idea of systematic proximity. Generally speaking, it is true to say that statistical comparisons of over-all measurements and indices are of the greatest value in assessing degrees of affinity in forms already known to be quite closely related—e.g., sub-species or geographical races, but they become of less and less practical value as the relationship becomes more remote and the types to be compared become more disparate.

Since the pioneer studies of Karl Pearson and his colleagues, the application of biometrics to taxonomic enquiries has become commonplace. But, because statistical methods are sometimes applied uncritically, and without due appreciation of the morphological and phylogenetic basis of taxonomy or of the fundamentals of the phenomena underlying the data to be measured, they have been open to criticism and are in serious danger of becoming discredited. For this reason it seems worth while drawing attention to a number of fallacies which are often overlooked by workers in this field, particularly in the field of palaeo-anthropology by those who are not morphologists by training or are not statistical experts.

1. THE FALLACY OF RELYING ON INADEQUATE STATISTICAL DATA

One of the limitations of the biometric analysis of taxonomic characters depends on the fact that, if adequate statistical methods are employed, the analysis of even a few measurements entails a very considerable amount of work. Consequently there is a danger of relying on too few measurements, a danger which is of course very seriously increased if these happen to have little taxonomic relevance. The comparison of such measurements may lead to the statement that (say) a fossil bone or tooth shows no significant difference from that of *Homo sapiens*, or perhaps from that of the Recent anthropoid apes. But clearly such a statement is of doubtful value (and may actually be very misleading) if at the same time account is not taken of other morphological features which may, in fact, be much more

* This paper represents the substance of a section which has been incorporated in "The Fossil Evidence of Human Evolution", published by the University of Chicago Press, 1955

relevant for assessing affinities. An example of this difficulty is provided by the famous case of *Hesperopithecus*. This generic name was given to a fossil tooth found in Nebraska in 1922, on the assumption that it represented an extinct type of anthropoid ape. Part of the evidence for this assumption was based on a comparison of the over-all measurements of the tooth with a series of ape teeth, for these metrical data established clearly that in this respect the fossil tooth falls within the range of variation shown in Recent apes. However, it was the critical eye of a comparative anatomist, with a long experience of the examination and discrimination of palaeontological material, which drew attention to certain "non-metrical" morphological details throwing serious doubt on the original interpretation. As is well known, the tooth proved later to be that of a fossil peccary. This example, which has a certain historical interest in the field of Primate palaeontology, is quoted here not in criticism of those who were responsible for the mistaken identification (there can be few palaeontologists who have not erred in this way at some time or another !), but to emphasize that two or three over-all measurements of a tooth can only express an insignificant proportion of all those metrical elements which contribute to its shape as a whole. This applies also, of course, to skulls or individual bones.

2. THE FALLACY OF TREATING ALL METRICAL DATA AS OF EQUAL TAXONOMIC VALUE

It has already been emphasized that morphological characters vary greatly in their significance for the assessment of affinities. Consequently it is of the utmost importance that, in applying statistical methods, particular attention should be given to those characters whose taxonomic relevance has been duly established by comparative anatomical and palaeontological studies. This *principle of taxonomic relevance* in the selection of characters for biometrical comparisons is one of great importance, but it is also rather liable to be overlooked. It may be asked how the distinction is to be made between morphological characters which are relevant or irrelevant for taxonomic purposes. The answer to this question is that each natural group of animals is defined (on the basis of data mainly derived from comparative anatomy and palaeontology) by a certain pattern of morphological characters which its members possess in common, and which have been found by the pragmatic test of experience to be sufficiently distinctive and consistent to distinguish its members from those of other related groups. The possession of this common morphological pattern is taken to indicate a community of origin (in the evolutionary sense) of all the members of the group, an assumption of which the justification is to be found in the history of palaeontological discovery. But, as a sort of fluctuating background to the common morphological pattern, there may be a number of characters, sometimes obviously adaptive, which not only vary widely within the group but overlap with similar variations in other groups. Such fluctuating characters may be of importance for distinguishing (say) one species from another within the limits of the family, but they may be of no value by themselves for distinguishing this family from related families. In other words they are taxonomically irrelevant so far as inter-familial relationships are concerned. The same applies to other major taxonomic categories such as super-families, sub-families, and so forth. For example, among the lemurs the over-all dimensions (length and breadth) of the molar teeth may provide useful criteria for distinguishing between the various species and sub-species of the Galaginae or between those of the Lorisinae, but they could not be expected to be of any value in differentiating between these two sub-families.

So far as the Hominidae are concerned, the principle of taxonomic relevance may be illustrated by reference to the extinct genus *Pithecanthropus*. The available evidence indicates that in this type the morphological features of the skull and jaws are very different from those of *Homo sapiens*, while the limb skeleton is hardly distinguishable. Clearly, therefore, if the question arises whether the remains of a fossil hominid are those of *Pithecanthropus* or *Homo sapiens*, for taxonomic purposes the morphological features of the skull and jaws are the relevant characters to which attention should be primarily directed. In the study of fossils representing early phases in the evolution of major adaptive radiations, their affinities must be determined by a study of those characters whose taxonomic relevance may be inferred from a consideration of the main trends of evolution as demonstrated by comparative anatomical studies and by extrapolation from the fossil record so far as the latter is available. For example, the initial evolutionary segregation of the Hominidae from the Pongidae was almost certainly dependent on modifications related to the development of an erect bipedal gait (see Washbrun, 1950, on this point). Hence, in assessing the affinities of the *earlier* representatives of the Hominidae (whose taxonomic position may be in some doubt), the skeletal characters of the pelvis and hind-limb are likely to be of much greater importance than those of the fore-limb. As the palaeontological record now shows, also, the morphological details of the dentition are likely to be of much greater taxonomic relevance than the actual over-all dimensions of the teeth and jaws, or the cranial capacity. As a further example of the principle of taxonomic relevance, we may refer to the dentition of some of the fossil representatives of the Pongidae. In these the incisor teeth are so similar to those of *Homo* (and even *Homo sapiens*) as hardly to be distinguishable. On the other hand, in all known pongids the canine teeth are quite different. Obviously, therefore, in determining as between the pongid or hominid affinities of a fossil hominoid, the canines have a much higher degree of taxonomic relevance than the incisors.

In order to keep within reasonable limits the number of measurements to be used for the statistical comparison of a fossil bone or tooth with related types, the rational procedure is first to make direct visual observations, selecting for comparison just those features which are known to have taxonomic value for the problem in hand. In many cases differences or resemblances may be so obtrusive as to obviate the need for statistical methods altogether. On the other hand, if differences and resemblances are not immediately apparent on visual inspection, special *ad hoc* measurements and indices may then be devised in order to test those characters which can reasonably be expected to be of value in the assessment of systematic affinities in any particular case. Only negative results are to be anticipated if routine measurements of little or no taxonomic value are employed.

3. THE FALLACY OF TREATING CHARACTERS SEPARATELY AND INDEPENDENTLY, INSTEAD OF IN COMBINATION

This fallacy has recently been treated in some detail by Bronowski and Long (1952). They point out that a bone or a tooth is a unit, and not a discrete assembly of independent measurements, and that to consider their measurements singly is likely to be both inconclusive and misleading. The right statistical method, they emphasize, must treat the set of variates as a single coherent matrix. This can be done by the technique of multivariate analysis, which is essentially a method (not possible with more elementary techniques) that can be used for comparing morphological *patterns*. In principle the application of the technique is straightforward enough, but it requires care and discrimination, a sound knowledge of morphology, and also a considerable experience of statistical methods. A number of measurements or indices of a bone or tooth are selected, which are judged on morphological grounds to be taxonomically significant, and from these the average, variances and correlations for a number of specimens are calculated. It is then possible to construct a numerical picture of the size and shape of the bone or tooth (and of the extent to which they vary), and to express this as a discriminant function. Such functions may be used for deciding whether (say) a fossil hominoid tooth is more likely to belong to a pongid or a hominid type, provided, of course, that the particular discriminant functions already calculated for the two families are sufficiently distinct. Bronowski and Long (1953) have emphasized the value of multivariate analysis by applying it to a controversial issue which had arisen in regard to certain teeth of the South African fossil genus, *Australopithecus*, and they were able to resolve the controversy by demonstrating very positively their hominid character.

4. THE FALLACY OF INADEQUATE OR INACCURATE STATISTICAL TREATMENT

This fallacy has been dealt with in part in the preceding section. The possibility of inaccuracies of computation is one which needs to be borne in mind, for cases have occurred in which such errors have led to rather serious mis-statements and misunderstanding. The amateur statistician needs to check and re-check his calculations so that there can be no possible doubt about the accuracy of his final figures. One of the disadvantages of scientific papers incorporating elaborate statistical analyses is that, since only the end results of the calculations are usually published, the latter cannot be checked by the reader. It is not a little disconcerting to contemplate the possibility that simple errors of calculation have occasionally occurred in the biometrical work of the non-professional statistician, leading to results, the falsity of which may not become apparent for some time.

5. THE PRINCIPLE OF MORPHOLOGICAL EQUIVALENCE IN MAKING STATISTICAL COMPARISONS

Failure to understand this principle is perhaps one of the most serious sources of fallacy likely to affect studies by those who are not thoroughly acquainted with the morphology of the skeletal elements with which they are dealing. A simple (but rather crude) example may be offered by referring to a measurement often employed in craniology—the auricular height. This is commonly taken by measuring the maximum height of the skull (in the Frankfurt plane) from the auditory aperture, and in comparing different racial groups of *Homo sapiens* it gives an index of the height of the brain case at this particular level. But, in comparing *Homo sapiens* with (say) the gorilla, it would clearly be misleading to employ the same technique, for in male gorillas the height of the skull is often considerably extended by the development of a powerful sagittal crest. If such a comparison were made, it would be a comparison of the height of the brain case in *Homo sapiens* with the height of the brain case *plus* a sagittal crest in the gorilla and would have no meaning from the morphological viewpoint. This is, of course, an extreme example, but it is perhaps not fully realized that similar (if less obvious) fallacies may be incurred in other craniometric work in which over-all measurements of the skull are commonly equated with each other. In comparing skulls of closely related groups such measurements may be sufficiently equivalent morphologically to make direct metrical comparisons valid. But if they are used to compare, say, a modern European skull with the skull of the fossil genus *Pithecanthropus*, quite serious difficulties are involved. For example, in the European the glabello-maximal length is an approximate measurement of the maximal length of the brain case. But in the *Pithecanthropus* skull it measures a good deal more, for the glabello-maximal length is complicated by the exaggerated development of a massive supra-orbital torus, the great thickness of the skull, and the projection backwards of an exaggerated occipital torus. The over-all glabello-maximal measurement is thus not strictly comparable (in the morphological sense) with that of a European skull—in both cases it involves a number of different elements and these may be independently variable among themselves. Again, in the European skull the maximal width is commonly situated in the parietal region, while in the *Pithecanthropus* skull it is situated in the temporal region. Thus to compare the maximal width in the two skulls is to compare measurements which also are not morphologically equivalent. In fairness to physical anthropologists generally, it must be stated that these sources of fallacy in comparative osteometric studies are usually quite well recognized, but this may not always be the case with the less careful workers. As a further example we may take the lower front premolar tooth in the Hominidae and Pongidae. One method which has been used for measuring the length of this tooth is to take the maximal antero-posterior diameter in the axis of the tooth row. But, as is well known, in the anthropoid apes the front lower premolar is commonly rotated on its vertical axis so that the axis corresponding morphologically to the transverse axis of the hominid premolar is directed obliquely postero-medially.

Thus, to compare the maximal antero-posterior diameters in the axis of the tooth row in the two groups is to compare dimensions of the premolar itself which again are not morphologically comparable. What is actually being compared in this case is the maximal antero-posterior *space* occupied by the premolar in the tooth row—a very different thing. In comparing over-all measurements of skeletal elements, it is of the greatest importance that the morphological basis of the dimensions to be compared should be stated very precisely indeed.* For the amateur biometrician, when comparing bones or teeth of different shapes, may fall into the trap of comparing dimensions which (because of the *different* shape) are not morphologically comparable, and then on the basis of this false comparison may conclude that because these dimensions are similar the bones or teeth are actually of the *same* shape. We may reiterate here (what has already been pointed out) that while biometrical studies of immediately related forms belonging to the same restricted group (such as a species or sub-species) may be expected to give fair comparisons which approximate sufficiently closely in their morphological equivalence, the statistical comparison of different genera which show a greater disparity of form needs to be carried out with a very critical appreciation of the technical difficulties involved.

6. THE FALLACY OF COMPARING SKELETAL ELEMENTS IN INDIVIDUALS OF DIFFERENT AGE, SEX AND SIZE

This is a fallacy which is perhaps hardly likely to occur in the hands of careful workers—and yet it does occur from time to time. It is well recognized that age changes may lead to quite considerable modifications in the structural details and proportions of skeletal elements. In the skull, for example, they are so marked that it would clearly be fallacious to compare a few measurements of the adult skull of a primitive hominid with those of juvenile skulls of an anthropoid ape and to infer from such a comparison that the former is not markedly different from anthropoid apes in general. In regard to sex differences, again, it would obviously be misleading to compare the dimensions of the canine teeth in fossil hominoids of presumably male sex with the relatively small teeth of a female gorilla, and to conclude therefrom that in these particular dimensions the fossil teeth fall within the range of variation of those of Recent apes. The sex variation needs to be taken into account in such a case, as it does also in comparisons of morphological and metrical features of the skull and skeleton in general.

The factor of body-size in statistical comparisons is perhaps of even greater importance, for the reason that it has been overlooked much more frequently. For example, differences of proportions in the skull and skeleton in Primates of different size may be merely an expression of allometric growth, or they may be related to the mechanical requirements dependent on differences in body weight. In either case, of course, they may be of very little taxonomic importance (except perhaps in the determination of specific or sub-specific distinctions). Thus, in quadrupedal mammals the relative thickness of the leg bones is a function of the absolute size of the animal, for the strength of a bone as a supporting structure varies as its cross-sectional area (*i.e.*, as the square of the linear dimensions of the animal) while the weight of the animal varies as the volume (*i.e.*, as the cube of the linear dimensions). In heavier mammals, therefore, the leg bones are relatively thick and their actual shape may thus be different from those of lightly-built (but still quite closely related) types. Clearly, then, it may be very misleading to compare the "robusticity index" of (say) the femur in Primates of very

* It may be argued that the fallacy of morphological non-equivalence must almost necessarily be involved in any over-all measurement of a skeletal structure, since such a measurement is bound to include a number of different components which may vary independently. For example, two skulls may show the same thickness of cranial wall, but the latter may actually be composed of different proportions of the outer table, inner table and diploe. This, of course, is perfectly true, but it only serves to emphasize still more strongly the need for care in stating the morphological fundamentals of the biometric data employed.

different body size (*e.g.*, hominids, apes and monkeys) and then assume degrees of affinity or divergence without any reference at all to the body size factor. The differences in shape of bones may be even more accentuated by the fact that in larger animals the muscular ridges, tuberosities and so forth are much more powerfully developed. Nor does this difficulty only apply to limb bones. In the skull it is well known that, in closely related animals account must be taken of the factor of allometry in comparing relative size of brain case and relative size of jaws, and differences in the proportion and indices of these structures may again be reflected in differences depending on the degree of development of muscular ridges, or of bony features which have developed in response to mechanical stresses. Thus, for example, it would be futile to compare cranial indices of a gorilla with those of a small monkey with any idea of drawing taxonomic conclusions, unless factors of body size are first taken into account. For the same reason (though at first sight the case is a less obvious one) it would be misleading to make a direct comparison of cranial indices in the small and delicately-built skull of a pygmy chimpanzee with those of a large and massive skull of a fossil *Australopithecus*. And, even in comparisons of modern human skulls of a single homogeneous series, account needs to be taken of absolute size, for it is well established by biometrical studies that there is a significant correlation between form (as expressed in cranial indices) and absolute size.

It can hardly be over-emphasized that in comparing dimensions and indices of skull, limb bones, pelvis or other skeletal elements of Primates generally, taxonomic conclusions must first be preceded by an enquiry into all the complicating factors related to body size, an enquiry which may need elaborate statistical studies and which most certainly requires an intimate knowledge of the structural responses of skeletal elements to functional demands.

7. THE FALLACY OF COMPARING MEASUREMENTS TAKEN BY DIFFERENT OBSERVERS USING DIFFERENT TECHNIQUES

The dangers of this fallacy have been emphasized again and again by biometricians, and the only excuse for mentioning it here is that it is still overlooked by some writers. If it is certain that the different observers are using identical techniques for recording their measurements the latter may be employed for comparative studies, but they still need to be used with the greatest circumspection (particularly in the case of very small objects where measurements need to be accurate to a fraction of a millimetre). Where it is apparent that different observers are not employing precisely the same techniques, statistical comparisons must necessarily be stultified. In anthropological craniometry attempts have been made (with some success) to secure general agreement on the definitions of points and planes which serve as a basis for statistical measurements (see, for example, Buxton and Morant, 1933). In regard to other elements of the skeleton, and also the dentition, the virtual absence of standardization of metrical technique renders comparisons between different observers very hazardous indeed.

8. THE FALLACY OF RELYING FOR THE ASSESSMENT OF AFFINITIES ON THE BIOMETRICAL ANALYSIS OF CHARACTERS WHICH MAY HAVE NO GENETIC BASIS

It is not always recognized that, during the period of growth, bone is a very plastic material. That is to say, its form may be readily modified by the mechanical effects of pressure and traction of the soft parts immediately related to it, and also by the effects of dietetic deficiencies or constitutional disturbances of one sort or another. This needs to be taken into account as a possible source of fallacy in the attempts which have sometimes been made to assess the affinities of the various racial groups of *Homo sapiens* on the basis of osteometric data, particularly in those cases where differences are so slight as only to be detected by statistical methods. So far as palaeo-anthropology is concerned, also, it is a factor which always needs to be taken into account when seeking evidence for the differentiation of geographical variants of the same general type. For example, it has been argued that the Javanese and Chinese representatives of the Pleistocene genus

Pithecanthropus are taxonomically distinct because (*inter alia*) the thigh bone of the latter shows a flattening of the shaft (*platymeria*) which is not present in the latter. But, apart from the fact that this is a feature which shows considerable variation within the limits of the species *Homo sapiens*, there is some suggestive evidence of an indirect nature that the degree of flattening of the shaft may depend on nutritional factors (Buxton 1938). If this is so, it is not a genetic character which can be properly used for taxonomic reference. How far nutritional or other post-natal influences may determine minor differences in cranial or facial proportions is still quite uncertain. It is for this reason, of course, that in the study of modern populations physical anthropologists are now placing less reliance on the comparisons of traditional anthropometry, and are concentrating their attention on characters whose genetic composition is directly ascertainable and by reference to which racial groups can be classified objectively on the basis of gene frequencies.

GENERAL CONSIDERATIONS

The enumeration in this paper of the difficulties and hazards involved in the application of statistical methods for assessing the affinities of fossil types is not, of course, intended as a disparagement in principle of the methods. On the contrary, it is precisely because their importance is well recognized that attention has been drawn to possible fallacies in their application. For, if these fallacies are not taken into full account, the inexperienced biometrician may easily bring into serious discredit the methods which he employs. Similarly, it is not intended to suggest that the hazards enumerated are not well understood by the professional statistician; but the point is that the literature of palaeo-anthropology makes it clear that they are not always recognized by the amateur statistician. It is a matter of outstanding importance to ensure that conclusions based on statistical methods are entirely valid, and to this end it is necessary to ensure that the methods themselves are adequate. Failure to do so has no doubt been mainly responsible for some of the inconclusive or misleading statements regarding morphological resemblances or differences which have been made as the result of statistical comparisons of fossil Primate material. For this reason the publications of Bronowski and Long, to which reference has been made, are of particular importance, for they have clearly shown that some of the most serious limitations of statistical analysis of morphological elements may be avoided by applying methods of multivariate analysis.* But they also emphasize that this method is by no means technically simple, and it also involves long and arduous computations. Finally, it should be stressed that the selection of the actual measurements used for statistical treatment requires the most careful discrimination, and the latter needs to be based on a thorough appreciation of the morphological fundamentals involved.

REFERENCES

BRONOWSKI, J., and LONG, W. M., 1952, Statistics of Discrimination in Anthropology. *Amer. Journ. Phys. Anth.*, Vol. 10, p. 385.

——, 1953, The Australopithecine Milk Canines, *Nature*, 172, 251.

BUXTON, L. H. D., 1938, Platymeria and Platycnemia. *Journ. Anat.*, Vol. 73, p. 31.

BUXTON, L. H. D., and MORANT, G. M., 1933, The Essential Craniological Technique. *Journ. Roy. Anth. Inst.*, Vol. 63, p. 19.

WASHBURN, S. L., 1950, The Analysis of Primate Evolution with particular reference to the Origin of Man. *Cold Spring Harbor Symposium on Quantitative Biology*, Vol. 15, p. 67.

MAHALANOBIS, P. C., MAZUMDAR, D. N., and RAO, C. R., 1949, Anthropometric survey of the United Province, 1941. *Sankhya*, Vol. 9, p. 89.

* It is of particular interest to note that the technique of multivariate analysis was successfully applied to anthropological problems some years ago by Mahalanobis, Majumdar and Rao (1949).