Multidimensional Scaling and the Analysis of Human Biological Diversity in Subsaharan Africa

G. P. RIGHTMIRE
Department of Anthropology, State University of New York, Binghamton, New York 13901

KEY WORDS Multidimensional scaling · Biological distance · Human variation · Africa.

ABSTRACT Much information concerning intergroup variation in Subsaharan Africa has been collected by Hiernaux, who has calculated measures of biological distance (Δg) between pairs of 101 African human populations. The resulting very large matrix of distances is not easily interpreted through simple inspection. Therefore non-metric multidimensional scaling is here applied in order to produce a configuration of 60 selected groups in a space of reduced dimensionality. A three dimensional solution shows acceptable stress, and this map of populations is used to support and extend Hiernaux's earlier conclusions.

In a detailed study of human diversity in the Subsahara, Hiernaux ('68a) presented an impressive compilation of serological, anthropometric and qualitative morphological information drawn from some 460 African populations. These data have been treated descriptively, in univariate fashion, and various correlations are provided with the aim of identifying associations of stature, blood groups and other traits with climate or geography. The study also attempted to relate populations on a generalized (multivariate) basis, to assess previous efforts at racial classification. For a set of populations on which at least six common characters could be measured or determined, Hiernaux has calculated Δg, a simple biological distance statistic which standardizes variable means or frequencies by dividing by sample ranges (Hiernaux, '65). The resulting matrix of distances between all pairs of 101 groups is used to demonstrate that no large clusters corresponding to traditional African racial or subracial divisions can be identified. Further analysis of these distances has subsequently been carried out in other papers (Hiernaux, '68b; '72), and the data are examined once more in the present contribution.

Despite difficulties with Δg itself (as here computed from both anthropometric means and blood group gene frequencies, with varying numbers of characters entering into the distance determinations), it is clear that Hiernaux's treatment of such a diverse assemblage of peoples is unique and valuable. Other surveys have been limited to relatively small numbers of African tribes, and large scale comparisons have not been carried out before. However, there are obvious problems encountered in this sort of continental approach to human variation, such as how best to deal with the very large numbers of distances obtained. The 5050 separate values of Δg given in Hiernaux's ('68a) original work vary from a low of seven to a maximum of more than 2,800 units, and this matrix contains much more information than can easily be digested through simple inspection. Without the assistance of ordination techniques now widely employed to summarize patterns concealed within such distance or similarity matrices, Hiernaux recognized 25 small "clusters," 23 of which consist of two groups only. The two remaining clusters involve three populations, and large constellations are not apparent. Further analysis of these distances was carried out along what Hiernaux refers to as non-taxonomic lines. For example, the most central populations (those having the lowest values of Δg, when distances to all other groups are averaged) were identified, and explanations for their positions explored. Several Bantu-speaking peoples show low average Δg values, attributed both to absence of marked adaptation to an extreme desert or forest environment and to genetic ties with an ancestral West African stock. This reasoning is based on consideration of only a small subset of the distances available, and such approaches may not provide as full interpretations of group interrelationships as can be obtained by other means.

In an effort to extract a maximum of information from the set of distances available, non-metric multidimensional scaling is here employed to produce a configuration of African populations in a space of reduced dimensionality.

METHODS

Various techniques for the analysis of interrelationships given in a matrix of similarities or distances are now available. Where this matrix contains character correlations, for example, principal components may be extracted and group positions projected onto these primary axes. If the data take the form of intergroup distances rather than correlation coefficients, then approaches such as multidimensional scaling (Shepard, '62a,b; Kruskal, '64a,b) or principal coordinates analysis (Gower, '66) should be preferable. Presumably either scaling or coordinates analysis could be employed successfully with Δg as input, and results obtained on diverse test data are reported to be similar (Rohlf, '72). However, Rohlf's work does suggest a preference for multidimensional scaling, though this technique may compress some shorter distances and stretch the larger ones in a manner not encountered in coordinates configurations. Differences between close groups are generally shown more accurately by the scaling routine, which is the one used here.

Given any set of distances between groups, multidimensional scaling can be used to find configurations of these groups in spaces of successively fewer dimensions. Program MDSCAL, part of the OSIRIS III package of programs distributed by the Institute for Social Research, University of Michigan, permits scaling in a maximum of ten dimensions and a minimum of two. The procedure is an iterative one, and within a space of specified dimensionality the groups are moved about to obtain a monotone relationship between the original proximity measures and the distances in the configuration. The extent of success in this endeavor is expressed as a measure of stress, as discussed by Kruskal ('64a,b). Stress is thus high when the fit of the new distances to the original rank order is poor and becomes lower as this fit improves. Zero stress implies a perfect monotone relationship between the two sets of coefficients. When a series of iterations has produced a configuration showing minimal stress in some number of dimensions, the procedure is terminated; spaces of fewer dimensions are then tried, and further configurations computed until programmed possibilities are exhausted. Generally, configurations of higher dimensionality show lower stress, while more restricted solutions fit less well. The decision as to which of several solutions is the most appropriate for study may thus be difficult.

Another problem encountered is the limitation imposed by MDSCAL on the number of groups included in the analysis. Unfortunately this must be kept to 60, and several of the populations contained in Hiernaux's final listing ('68a: 107) have had to be discarded. Selection of 60 groups for scaling has been made so as to retain as much as possible of the geographic, ethnic and linguistic variation represented in the original set of 101 populations.

RESULTS AND DISCUSSION

Multidimensional scaling in ten dimensions yields a configuration showing low (4.3%) stress, as expected, but such a solution is hardly open to ready interpretation. Trials specifying successively fewer dimensions give configurations of higher stress, and the final, two dimensional array exhibits relatively poor fit (15.5% stress) to the set of original Δg values. If this two dimensional map is rejected, as presenting a simple yet distorted impression of intergroup relationships, then some one of the more complex solutions must be used instead, and the choice is not obvious. One approach to this question recommended by Kruskal ('71) is to compare stress against dimensionality for a number of trials and to look for discontinuities in the resulting curve. If the curve changes direction sharply at some point, then the "true" dimensionality of the data can be specified. When this is done, the plot shows a minor elbow at three dimensions and another at four, while for larger numbers of dimensions the curve is relatively flat. This procedure thus provides no firm answers but suggests that either the three or four dimensional solutions may be appropriate. The latter is associated with stress of 8.5% but has the same drawbacks as do all configurations of higher dimensionality: complexity leading to difficulty with interpretation. The three dimensional map shows higher stress (11.3%, which rates as about "fair" on the scale supplied by Kruskal, '64a) but is more easily translated into biological terms, and probably the advantages of dealing with this simpler configuration outweigh the resulting loss in accuracy of fit. Positions of 60 African tribes for this trial are shown in figures 1 and 2.

This analysis demonstrates that the groups do indeed form a large, semispherical cluster, as Hiernaux suggested. This cluster is relatively dense centrally, whereas points toward the periphery are more dispersed. Several of these peripheral groups seem to form loose associations among themselves, while others are so distant from all of their neighbors as to be termed isolates.

By and large, the more centrally placed populations are those which speak a Bantu dialect, though a diversity of geographic areas are represented. This grouping includes the Ewondo and Basa of southern Cameroun, the Nyoro of Uganda, the Haya and Nyamwezi of Tanzania and others which exhibit relatively low average values of Δg. Some other Bantu-speakers such as the Humu (or Amba) and several West African peoples are peripheral in the plane of dimensions I and II, while the Bamum and Bushong are extreme on axis III. The Mbundu of Angola are also Bantu-speaking, though obviously far removed from all of their linguistic associates. Why this one southwestern tribe should occupy an isolated position is unclear, particularly as the Kwanyama of the same area and similar cultural inventory are well within the central cluster.

It is also apparent from the figures that other groups in addition to the bulk of Bantu-speaking peoples are fairly centrally located in the configuration. The Dyola of Senegal, for example, speak a West Atlantic language which belongs to the same large (Niger-Congo) family as do all of the Bantu dialects (Greenberg, '63). The Koniagi, Badyaranke, Balante, Kasena and Bassari are also Niger-Congo speakers of West Africa, and these several non-Bantu groups are closely associated in the plots, though they do not form a separate cluster. If Hiernaux's ('68b) reasoning is correct, a number of these most central populations sharing broad linguistic affinity may represent an ancient African stock, perhaps that from which both Bantu and other Niger-Congo speakers have since been derived.

However, the Alur, Luo and Kakwa, located toward the periphery of the configuration as seen in figure 1 but still part of this central assemblage, are not only linguistically non-Bantu but also fall outside of the entire Niger-Congo speaking family. These peoples of Kenya and Zaïre are Nilotic Negroes, and their language is classed within the Eastern Sudanic division of Chari-Nile (Greenberg, '63). Despite this linguistic difference, these particular tribes are claimed by Hiernaux to be biologically close to the same West Central African stock located in or slightly north of the forests which has since produced the Bantu and other Niger-Congo speaking peoples already referred to. In sum, Hiernaux's analysis designates some Nilotes, some West African Atlantic and Voltaic speakers, and a number of Bantu-speaking Negro tribes as the biological descendants of a nuclear population or set of closely related populations occupying West Central Africa several millenia ago. Expansion of this nuclear population and movement especially of the Bantu-speakers have resulted in dispersal of tribes into central and southern Africa, toward the areas they occupy today.

This is a large hypothesis and one which rests as much on linguistics as on the biological evidence supplied by Δg. The latter is distilled and summarized by means of scaling, and it is clear that the central aggregate of groups does comprise culturally and linguistically diverse peoples. Despite the peripheral positions of some Mande and Kwa-speakers (Gio, Kru and Grebo) and some Nilotes (Nuer, Masai), it should be emphasized that the Bantu-speakers do not uniformly segregate themselves from other Negro Africans. This

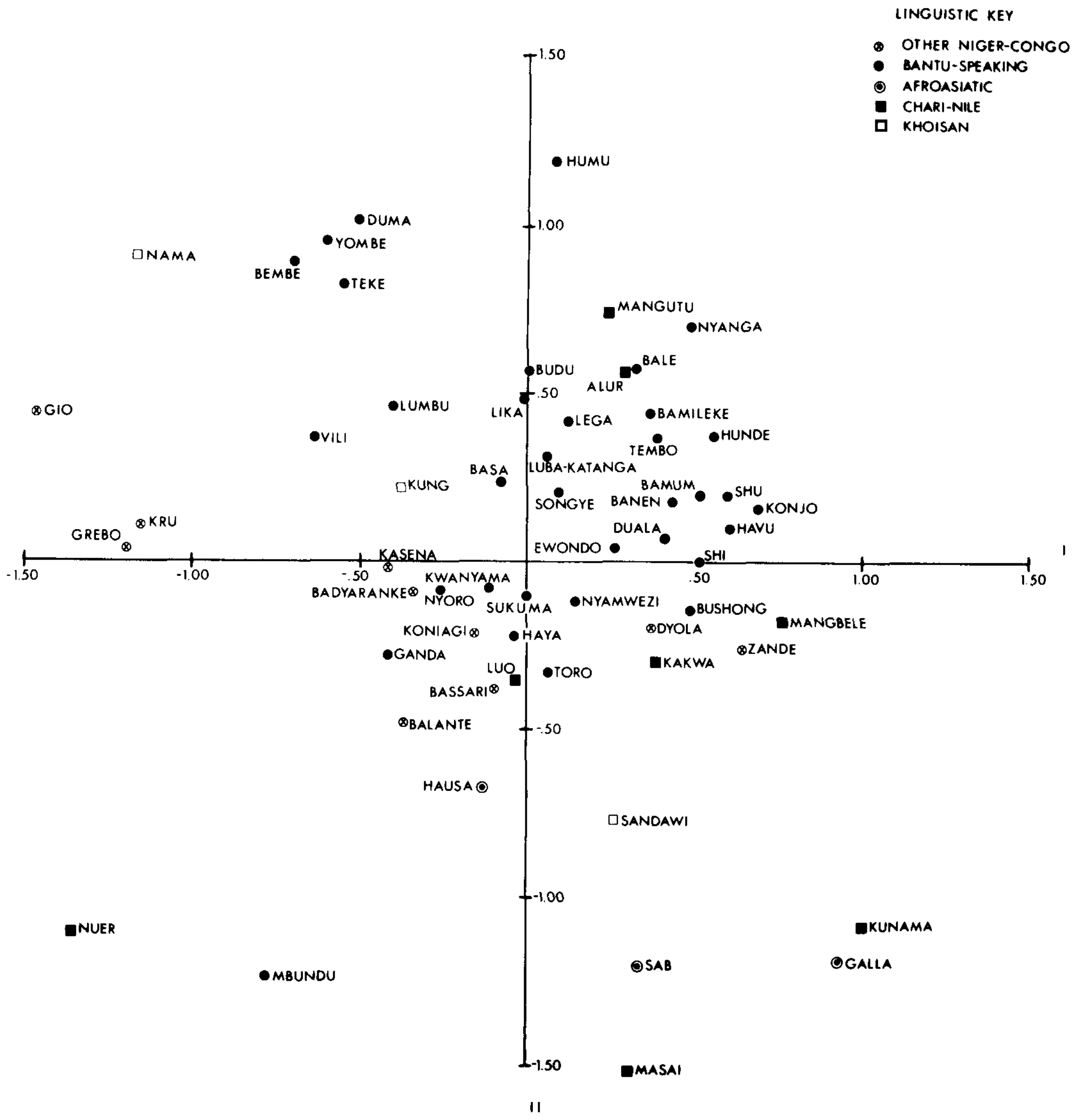

Fig. 1 Results of non-metric multidimensional scaling of Δg values for selected African populations. Groups are located on the first two dimensions of a three dimensional configuration showing stress of 11.3%. Group names are those employed by Hiernaux ('68a), while assignments to language families are based on Greenberg ('63). A linguistic key is provided primarily to facilitate identification of Bantu-speaking peoples and to demonstrate that there is little clustering corresponding to broad linguistic categories. This plot should be used with fig. 2 in order to appreciate intergroup relationships in all three dimensions discussed in the text.

evidence, however limited, reaffirms the need to drop Bantu as a term of biological significance, as has been urged by Tobias ('71) as well as by Hiernaux and others. Language and biological affinity simply need not keep in step, and numerous cases of discordance are on record (e.g. Spuhler, '72).

Another point, apart from difficulties with linguistic classifications, concerns

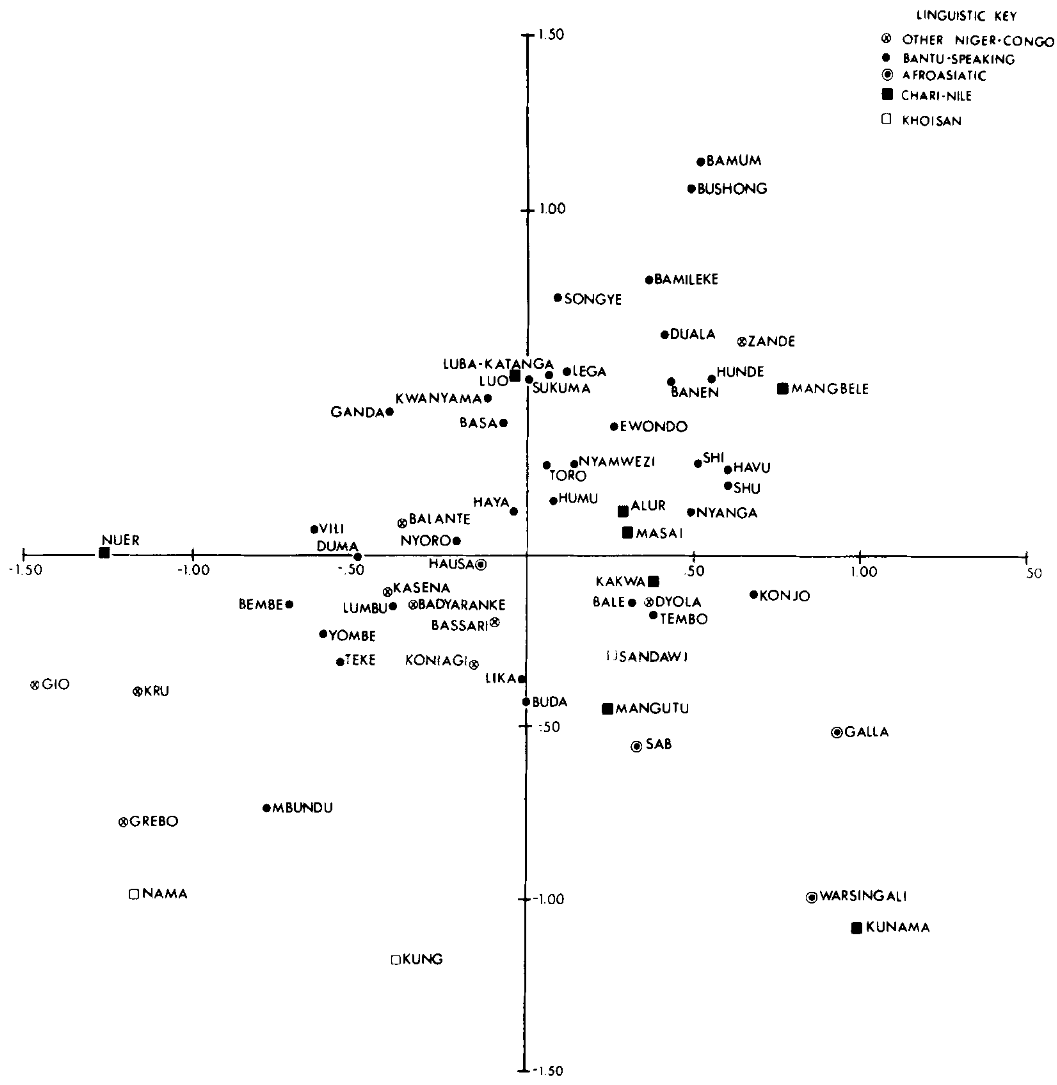

Fig. 2 Results of scaling of Δg values for 60 selected African populations. Groups are located on the first and third dimensions of a three dimensional configuration showing stress of 11.3%. This plot should be used with figure 1 in order to appreciate intergroup relationships in all three dimensions discussed in the text.

geography. For a long time it has been customary to subdivide the peoples of Africa into a series of local races. These groupings are traditionally defined in geographic as well as purely biological terms, and a principal aim of Hiernaux's project has been to discover whether such large clusters of populations would emerge from the analysis of a comprehensive set of distance measurements. In fact, discrete aggregates of tribes corresponding to broad geographic areas do not appear, as is evident from the figures. The northeastern peoples of the Horn (Sab, Galla, Warsingali and Kunama) do occupy an otherwise underpopulated section of the reduced space, and these groups form part of what Garn ('65) has called the East African local race. Such linear and often lighter-skinned inhabitants of Ethiopia and surrounding territory are usually presumed to have mixed extensively with Mediterranean Caucasoids and therefore to be less Negroid in genetic make-up than other Subsaharan Africans. This departure from "typical" African frequencies has been confirmed for various marker genes among Ethiopians (Ikin and Mourant, '62; Harrison et al., '69), so that the present peripheral position of the Galla and related Cushitic speaking groups is not surprising. That the Kunama should be so clearly associated with this cluster is worth noting, however, in view of the recent work of Mourant et al. ('74), suggesting that this Chari-Nile speaking population is much more Negro-like than other Ethiopian peoples so far tested.

Northeastern Africans aside, other populations refuse to sort themselves by geographic region. There is no clear segregation of a Nilotic or Sudanese local race, nor is there any tendency toward delineation of a large "true" Negro or "forest" Negro cluster encompassing tribes of West and Central Africa; and failure to document a separate Bantu-speaking subrace in southern Africa has already been mentioned. Probably, despite local variation, all of these populations of the Subsahara are best collectively described as African Negroes, and there is growing evidence that they share in a complex of serological and other traits that is essentially unique (see Tobias, '72 for a review).

There is finally the question of Bushman and Hottentot associations, and here the evidence is really not clearcut. In the plane of dimensions I and II, these two southern African groups are not obviously set apart from the large central concentration of Bantu-speakers and other Negroes, and there is substantial distance between them (more, for example, than separates the Kung Bushmen from the Bantu-speaking Lumbu). They are less isolated than the Nilotic Nuer, the Ethiopians, or even some Niger-Congo speakers of West Africa (e.g. Gio, Grebo and Kru, who appear relatively close to the Kung and the Nama in this two dimensional array). However, when dimension III is taken into account, the Bushmen are seen to be extreme in a negative direction, having the lowest position of all peoples included in the analysis. This third axis effectively removes the Kung and also the Nama Hottentots from the vicinity of their many Negro neighbors in the plane of dimensions I and II, and only the Liberian Grebo remain in some proximity. Other tribes which appear to be close to the Khoisan speakers in figure 2 are in fact widely displaced from them when all three dimensions are considered. If special ties with the Grebo are discounted, then Bushmen and Hottentots emerge as clearly peripheral; this result is not noted by Hiernaux but fits well with other studies such as those of Rightmire ('70) and Tobias ('72).

More than this is difficult to say on the basis of information contained in the battery of distances. Essentially, Hiernaux's conclusions regarding a relatively dense central swarm of linguistically diverse Africans are verified by multidimensional scaling. But in clarifying the associations of certain northeastern peoples on the periphery of this cluster, in identifying Bushmen and Hottentots as distinct from various Negro populations, and in otherwise providing a succinct summary of intergroup relationships, this exercise has extended the earlier analysis done without techniques of ordination.

ACKNOWLEDGMENTS

For computing assistance and for preparation of the figures, I am grateful to Linda Huff.

LITERATURE CITED

Garn, S. M. 1965 Human Races. Second ed. Thomas, Springfield.

Gower, J. C. 1966 Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika, 53: 325–338.

Greenberg, J. H. 1963 The Languages of Africa. Indiana University, Bloomington.

Harrison, G. A., C. F. Küchemann, M. Moore, A. J. Boyce, T. Baju, A. E. Mourant, M. J. Godber, B. G. Glasgow, A. C. Kopeć, D. Tills and E. J. Clegg 1969 The effects of altitudinal variation in Ethiopian populations. Phil. Trans. Roy. Soc., B, 256: 147–182.

Hiernaux, J. 1965 Une nouvelle mesure de distance anthropologique entre populations, utilisant simultanément des fréquences géniques, des pourcentages de traits descriptifs et des moyennes métriques. Comptes Rendus Acad. Sci., Paris, 260: 1748–1750.

——— 1968a La Diversité Humaine en Afrique Subsaharienne: Recherches Biologiques. Université Libre de Bruxelles.

——— 1968b Bantu expansion: the evidence from physical anthropology confronted with linguistic and archaeological evidence. J. Afr. Hist., 9: 505–515.

——— 1972 The analysis of multivariate biological distances between human populations: principles and applications to sub-Saharan Africa. In: The Assessment of Population Affinities in Man. J. S. Weiner and J. Huizinga, eds. Clarendon Press, Oxford, pp. 96–114.

Ikin, E. W., and A. E. Mourant 1962 A survey of some genetical characters in Ethiopian tribes. V. The blood groups of the Tigré, Billen, Amhara and other Ethiopian populations. Am. J. Phys. Anthropology, *20*: 183–189.

Kruskal, J. B. 1964a Multi-dimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, *29*: 1–27.

——— 1964b Non-metric multi-dimensional scaling: a numerical method. Psychometrika, *29*: 115–129.

——— 1971 Multi-dimensional scaling in archaeology: time is not the only dimension. In: Mathematics in the Archaeological and Historical Sciences. F. R. Hodson, D. G. Kendall and P. Tautu, eds.. Aldine, Chicago, pp. 119–132.

Mourant, A. E., A. C. Kopeć, E. W. Ikin, H. Lehmann, P. Bowen-Simpkins, I. L. C. Fergusson, M. D. Hellier, R. D. Jones and I. A. M. Roberts 1974 The blood groups and haemoglobins of the Kunama and Baria of Eritrea, Ethiopia. Ann. Hum. Biol., *1*: 383–392.

Rightmire, G. P. 1970 Bushman, Hottentot and South African Negro crania studied by distance and discrimination. Am. J. Phys. Anthrop., *33*: 169–195.

Rohlf, F. J. 1972 An empirical comparison of three ordination techniques in numerical taxonomy. Syst. Zool., *21*: 271–280.

Shepard, R. N. 1962a The analysis of proximities: multidimensional scaling with an unknown distance function. I. Psychometrika, *27*: 125–139.

——— 1962b The analysis of proximities: multidimensional scaling with an unknown distance function. II. Psychometrika, *27*: 219–246.

Spuhler, J. M. 1972 Genetic, linguistic and geographical distances in native North America. In: The Assessment of Population Affinities in Man. J. S. Weiner and J. Huizinga, eds. Clarendon Press, Oxford, pp. 72–95.

Tobias, P. V. 1971 The biological invalidity of the term Bantu. S. Afr. J. Sci., *67*: 517–520.

——— 1972 Recent human biological studies in southern Africa, with special reference to Negroes and Khoisans. Trans. Roy. Soc. S. Afr., *40*: 109–133.