

A Critical Examination of Numerical Taxonomy

Having completed our discussion of the various aspects of taxonomic procedure in biology from the point of view of numerical taxonomy, we now look to the task of critically evaluating the alleged and real weaknesses of the philosophy and methodology of numerical taxonomic procedure. Of course we wish also to stress positive contributions that numerical taxonomy has made and is making to biology, and the promise it holds for the future. In the first section we shall review briefly some of the more recent criticisms that have been leveled against one aspect or other of such work. In the second section we summarize unsolved problems of numerical taxonomy: Finally in Section 10.3 we shall point out how numerical taxonomy has led to a reevaluation of taxonomic principles and is leading to new knowledge of taxonomic facts as well as of evolutionary and ecological principles.

10.1 CRITICISMS OF NUMERICAL TAXONOMY

Ever since numerical taxonomy was first presented to the scientific community it has been subjected to criticisms on a variety of grounds. These criticisms have on occasion led to acrimonious and emotional controversy, much of which has fortunately subsided in recent years. It is important, however, for the reader of this book to be acquainted with a representative sample of the criticisms aimed at the

methods and philosophy of numerical taxonomy and we have therefore already discussed a variety of views contrary to ours at several places throughout the text. In this section we shall mention some fundamental criticisms that may not have been adequately covered earlier, many of which are put forth in several comprehensive critical reviews of numerical taxonomy to which we shall refer.

Before we embark on this discussion it is important for the reader in the early 1970's to have some historical perspective regarding the development of our views and of the ensuing controversy. When numerical taxonomy was first propounded in the middle 1950's, a well-developed and relatively stable theory of systematics was held by most biologists over the world. These views reflected the maturing of the important conceptual contributions that had been made during the 1930's and 1940's by the "New Systematics," an evolutionary view of biological systematics based on an integration of genetic and biosystematic knowledge. This impressive body of knowledge had resulted from the work of numerous outstanding scientists such as J. S. Huxley, E. Mayr, T. Dobzhansky, G. G. Simpson, G. L. Stebbins, Jr., B. Rensch, and E. Anderson.

The development at that time of numerical and phenetic taxonomy was not a sudden event, but a gradually increasing questioning of some of the established assumptions and techniques. Nor were the views we have propounded in the present book arrived at suddenly; they were, of course, developed gradually over a period of time and were modified in response to constructive criticism received from our colleagues. To illustrate the development, we may outline the following steps in the evolution of present day numerical taxonomy: (1) the discovery (or in historical perspective, *rediscovery*) that the measurement of similarity could be approximated (Sneath, 1957a; Michener and Sokal, 1957); (2) the realization that such a procedure in taxonomy was contrary to established philosophical tenets (e.g., Inger, 1958; Sokal, 1959); (3) the discovery (*rediscovery* may again be the more appropriate term) of the looseness, possible circularity, and nonoperational nature of most of the concepts of phylogenetic systematics (Sneath, 1961; Sokal, 1962b; Sneath and Sokal, 1962); (4) the further realization that purported and actual conventional practice in taxonomy were widely disparate (e.g., Michener, 1963; Sokal, 1964a); (5) the redefinition of taxonomic principles in terms of an empirical and operational science (Sokal and Sneath, 1963; Sokal and Camin, 1965); (6) the initiation of the heuristic phase in which the new insights gained through phenetic taxonomy were applied toward the development of new theory (Ehrlich and Holm, 1962; Sokal and Sneath, 1963).

The controversy surrounding numerical taxonomy increased subsequent to the publication of the *Principles of Numerical Taxonomy* (Sokal and Sneath, 1963) and is reflected in the records of numerous symposia held during the first half of the 1960's. The critical views were summarized by Mayr (1965) in an important article, the main arguments of which were answered by Sokal et al., (1965). Because these publications are widely known and accessible and also because of changes in

the points of view adopted by numerical taxonomists in the time span from the middle 1950's to the beginning of the 1970's, some of the early criticisms and controversies are no longer relevant today and are therefore omitted in this book. Persons interested in these historical aspects are referred to Sokal and Sneath (1963) and the cited papers, where they are discussed in detail. We should point out, however, that possibly because of linguistic barriers, arguments that had been debated and resolved, and that to some degree are no longer even relevant to modern numerical taxonomy, continued to be debated by some, e.g., Janetschek (1967) and Ziswiler (1967).

Even among some English speaking authors there remain misunderstandings of important points of our philosophy of classification. The book by Crowson (1970) is an example. Our views on unit characters, equal numbers of characters for all OTU's, and the implications of R studies are all misinterpreted. Another curious trend is illustrated by Crowson's book. No more than five references to numerical taxonomy are included in a bibliography of over 220 papers. Only one of these papers is dated 1967, the others are from 1963 or earlier. Many other important recent papers on taxonomic philosophy and methodology are not considered. It would be unfair to single out Crowson's book in this respect when some other recent works on systematics have been similarly deficient. Yet we should deplore a tendency among systematists to think of their field as unchanged and unchanging. What modern textbook of ecology or evolution (not to mention fields such as molecular genetics or neurophysiology) could appear in 1970 and not discuss the important concepts and contributions engendered during the 1960's?

Phyletics versus Phenetics in Taxonomy

This remains a matter for continued dispute in taxonomic circles. We have already discussed this topic in some depth in Section 2.6. Some authors dismiss phenetic taxonomy out of hand; Janetschek (1967) and Gisin (1964) apparently do not consider the possibility of a nonphyletic approach to taxonomy worthy of discussion. Ghiselin (1966, 1969) criticizes the concept of overall similarity severely. Whether similarity is "over all" or "over some" has already been discussed earlier (Section 3.8). But the fact that numerical taxonomists are able to measure phenetic-similarity and to do so quite successfully should be obvious even to the casual reader of the literature in systematics. Ghiselin's numerous other criticisms in these philosophically obscure papers have been answered sufficiently by Hull (1969) and Farris (1967c). L. A. S. Johnson (1968), in what is undoubtedly the most incisive and knowledgeable critical review of numerical taxonomy published to date, states that "... floral adaptations for pollination by long-tongued insects or by birds cannot have occurred in geological periods before such animals existed. Any phenetic classification which grouped organisms in a manner inconsistent with such a fact (taking all other relevant information into account) would not be

acceptable as even roughly consistent with phylogeny, and there is no reason why we should be asked to accept it simply because of claims of repeatability, objectivity, precision, or stability." The key point here is that one would not ask a systematist to accept a phenetic classification as the true phylogeny of the organisms. But one would ask him to accept it as the *phenetic* classification. Perhaps the crux of the matter is: Is Johnson prepared to accept a phenetic classification? He does not seem to recognize the general usefulness of phenetic classifications although he does not furnish cogent arguments why only phylogenetic classifications should be employed in taxonomy. This is an argument scarcely heard outside biology and not always very coherent within it: Constance (1964) in a revealing footnote, comments that many biologists appear to use the term "phyletic" to mean phenetic.

It is easy for phylogenetically oriented taxonomists to fall into the error of expecting phenetic taxonomies to yield the phylogenetic conclusions that they anticipate. As an example we may cite Kendrick and Weresub (1966), who, when classifying orders of basidiomycetes, obtained a phenetic classification that did not agree with the traditional orders. They reject numerical taxonomy on that account because implicit in their argument (although not clearly expressed) is that they wish a phylogenetic (possibly cladistic) classification. However, it should be obvious that a phenetic classification will not necessarily give them a cladogram or even a phylogenetic tree, whatever that may mean. Nor do they state anywhere on what evidence, other than their opinions as taxonomists, they base the correctness of the established classification. The reasons why the phenetic classification was so unsatisfactory have not been explained, although convergence may have been a cause. It is questionable whether the traditional higher categories in this group are as solidly established as Kendrick and Weresub seem to imply.

Another case in point is two statements by Cracraft (1967) discussing the merits of phylogenetic versus operational homology: "... rather it suggests that a non-evolutionary approach will fail to give us as accurate a picture of historical (biological) events as the evolutionary definition will." This statement is almost tautological. Obviously, a nonevolutionary approach need not provide an accurate picture of past events. Another quotation from the same author: "... there is the danger of both methods calling homologous structures (in the sense of common ancestry) that have changed radically with time nonhomologous. Furthermore, structures exist that are almost identical morphologically but that cannot be traced back to a common ancestor; here either method might incorrectly reach a decision of homology." Since operational homology is simply based on identity or similarity among characters and their relationships, the question of deciding whether these data are homologous in the sense of common ancestry and possible errors in such decisions are simply not relevant.

L. A. S. Johnson (1968) seems to feel that homologies must be worked out phylogenetically before any phenetic comparison can be useful (although he does

not show us how such phylogenetic analyses could be done). He does not think much of attempts at operational homology published up to the time of his own publication. However, as we have seen in Section 3.4 the work of Jardine and others has made impressive progress here. Also the evidence from falsifying homologies (Fisher and Rohlf, 1969) and from the pseudoscanning experiment (Rohlf and Sokal, 1967) indicate that the reliability of similarity estimates based on characters of dubious homology is much better than one would anticipate.

Mayr (1969a, p. 82) realizes that both cladistic and phenetic elements must enter taxonomic theory. Yet his fundamental contradiction is the statement "it is therefore completely legitimate to define taxonomic category in evolutionary (largely phylogenetic) terms, but to use evidence (comparative character analysis) that, as such, is almost entirely nonphylogenetic." He claims (p. 208) that a theoretical weakness of the purely phenetic method is its "inability to distinguish between phena and taxa." This is not true by the definitions that Mayr gives for phenon and taxon, since according to him a taxon is a matter of opinion (p. 4). It would be very difficult for any taxonomist, conventional or numerical, to decide how to turn one into the other but we would maintain that if anyone could do it, a numerical taxonomist could.

Criticisms of the Philosophical Bases of Numerical Taxonomy

Mayr's (1969a) discussion of the philosophical bases of classification suffers from overemphasis on categorizing ideas and from an eclectic use of citations to support his views. He attacks (1) essentialism and typology in their old fashioned manifestations, (2) nominalism, or the point of view that only individuals exist and all universals are artifacts of the human mind, (3) empiricism, and (4) cladism. He advocates evolutionary classification in which "the biologist classifies populations, not individuals or phena." The lower taxa are not "arbitrary aggregates, but reproductive communities tied together by courtship responses and separated from other similar units not by arbitrary decisions of the classifier but by isolating mechanisms encoded in the genetic program of the organism. . . The higher taxa, likewise, are characterized by the joint possession of components of an ancestral genetic program. . . Organisms have another unique property which distinguishes them from inanimate objects: they have a phenotype and a genotype. . ." Such a categorization into "good" and "bad" schools of taxonomy misses the important point that rather than conform rigidly to an orthodox taxonomic philosophy the new taxonomy must borrow and combine ideas from all. We have already rejected old-fashioned idealistic taxonomy (see Section 3.4) but a statistical typology (Sokal, 1962b) that seeks to discover the nature of patterns in the system of nature would surely be one of the goals of taxonomy. Mayr himself (1969a, p. 76) in describing the goals of his evolutionary classification says that "the taxonomist

no longer 'makes' taxa, he becomes a 'discoverer' of groups made by evolution." Yet this discovery must be based on phenetic evidence.

Although an exclusively nominalist point of view would be extreme, it seems to us desirable to proceed cautiously from entities that we *can* hope to define (individuals) to classes that we *may* hope to define. Mayr feels that numerical taxonomists err in their misinterpretation of causal relation between similarity and relationship. As Sokal (1969) has already pointed out, Mayr uses a poorly chosen example to illustrate his case, for he states (p. 68) that "it is exactly as with identical twins; two brothers are not identical twins because they are similar, but they are similar because they are both derived from a single zygote, that is, because they are identical twins." Mayr misses the point that the statement about the two being identical twins is a hypothesis derived from our observation of the very great similarity between any two individuals and our knowledge of genetics and cytology. On this point Mayr enlists Hull's (1967) statement of the view that phylogenetic inferences and classificatory procedure are not circular, but Hull's ensuing remark that most inferences of this sort are unwarranted is not quoted. Whatever the philosophical merits of empiricism it should be evident from the account in the present volume and in the by now very large literature of numerical taxonomy that numerical methods do work and taxa do "emerge automatically."

In summary the basic difficulty between our philosophy and the evolutionary classification of Mayr is that we try to distinguish between the process of classifying and the subsequent process of generating hypotheses about evolution, and Mayr wishes to combine these two. Yet in his entire volume he fails to tell us how to make such a combination. Our dissatisfaction with vague philosophies and methods has led to the principles and procedures propounded in our book.

The ideas by Gisin (1964) critical of numerical taxonomy (quoted approvingly by Mayr, 1969a) are based on quite obvious misreadings of the aims and methods of numerical taxonomy. Thus he does not, for example, understand that numerical taxonomists do study character correlations (R matrices). His later writings aiming at a quantum theory of evolution grope towards a synthetic view of cladistic and phenetic systematics, but in his later paper (Gisin, 1967) before his untimely death he had not as yet provided us with an operational method for achieving these goals. There seem to us two flaws in Gisin's system. The first is the lack of a quantitative approach to a task that must essentially be mathematical (although in his 1967 paper he begins to approach this goal). The second is the assumption that there are in fact distinct nonarbitrary categories such as species, genus, family, order, class, and phylum that, being there, must therefore be definable. Thus much of his theory (Gisin, 1966) is concerned with finding objective ways of defining these categories.

Criticisms of Methodology

Some writers feel that biological classification is so different from that of inanimate objects that few if any general principles and methods may be derived from a study of the latter. Mayr (1969a), surely no vitalist, still maintains a strong emphasis on the difference in principle between classifying inanimate objects and organisms. Burt (1966), Rollins (1965a), and Gisin (1964) are others who claim that the classifying of living organisms is not necessarily ruled by the logic appropriate to the classification of inanimate objects. We feel that the areas of common ground in classificatory theory are far greater than these critics allow.

An important point stressed by Mayr (1969a, pp. 205, 208) is whether the chosen sample of characters on which a similarity coefficient is based is sufficient for making a classification. He cites sibling species and sexually dimorphic groups (birds of paradise) as two phenetic extremes. In the former, phenetic similarity would underestimate their "drastic genetic differences"; in the latter, phenetics would exaggerate the differences. The pheneticist's answer to this problem is that (1) if overall phenetic differences correspond to the assumptions implied by these extremes, then this would be an important biological fact worth recording and measuring and would be useful in the generation of ecological hypotheses having to do with width of niches, evolutionary rates, and similar phenomena; (2) we really do not have any hard data measuring the similarity of either sibling species or highly sexually dimorphic species by appropriate methods of phenetic analysis and until such measurements are made, a statement such as Mayr's is clearly misleading. The criticism of equal weighting of characters as not conforming to evolutionary patterns (Mayr, 1969a, p. 209) would be valid if evolutionary patterns were the primary aim of phenetic taxonomy. Numerical phyletics has of necessity begun to weight characters (see Farris, 1969b).

Other criticisms illustrate the difficulty of thinking in quantitative terms even when discussing an inherently quantitative subject. An example is provided by the recent critique by Inglis (1970) on the nexus, nonspecificity, and matches asymptote hypotheses. Such hypotheses are not simply true or false: they hold in some places to a varying degree. We need to discover where and how far they hold, and even if they did not hold at all, it would not destroy the bases of numerical phenetics as Inglis implies. It would simply mean that our material was more difficult than we had thought.

It is easy today to point out the weaknesses of the mathematical and theoretical assumptions of the earliest papers in the field. L. A. S. Johnson (1968) is quite correct when he states that much more work in various areas of mathematics, especially topology, is needed to gain further insight into the problems of taxonomy. Several numerical taxonomists have also pointed out these problems (among others, Williams and Dale, 1965; Estabrook, 1966; Gower, 1967a; Jardine, Jardine, and

Sibson, 1967; Jardine, 1969b; Jardine, van Rijsbergen, and Jardine, 1969; Lerman, 1970; Jardine and Sibson, 1971), and have begun to attack them imaginatively. We must agree with L. A. S. Johnson (1968) that there probably is no one optimal classification, but his dictum that taxonomy is necessarily inexact and that compromise solutions are what we should settle for is surely too pessimistic. We may also wish to designate inherent principles of optimality that are mathematically defensible rather than those that imitate human psychological processes.

Burt (1966) and numerous others mention the importance of experience in taxonomy, but this can be countered by the work of Sokal and Rohlf (1970) as well as by the pseudoscanning study of Rohlf and Sokal (1967). In addition, we know of numerous unpublished studies by our students and others, who in term projects in numerical taxonomy courses have come up with acceptable classifications over a wide spectrum of organized nature, often with minimal (or no prior) knowledge of the particular group being classified.

Restrictions of Numerical Taxonomy to Special Data Sets

Some authors, though granting a measure of merit to numerical taxonomic techniques, would wish to restrict their application to special kinds of organisms or special ranks only, although no obvious limitations of this kind are evident in the existing literature (see Appendix A). Oldroyd (1966) feels that numerical taxonomy should be used exclusively at the species level but not above it. Throckmorton (1965) expresses similar views. Yet others, such as Mayr (1969a) have felt very strongly that the species, more than any other category, should be defined by biological rather than phenetic criteria. It seems difficult to us to find any logical grounds for restricting numerical phenetic studies to any one categorical level. Sokal and Crovello (1970) have recently emphasized that the supposedly biological criteria for species are dependent on phenetic considerations. And even though modern evolutionary theory uses the population as its basic unit, there is considerable difficulty in defining such a concept, as Sokal and Crovello (1970) point out at the end of their critique of the biological species concept.

Burt, Hedge, and Stevens (1970) state that computer taxonomy will make feasible "an attack on large compact groups of world-wide distribution which have just never been tackled as a whole." Michener (1963) maintains that numerical taxonomy should not be used in an exploratory way for a group of organisms to be classified for the first time, but should be used for refined analysis of a well established group, i.e., revision of a genus from the point of view of efficient use of a researcher's time. There have been so far few attempts at making actual time and efficiency estimates in classification and this question cannot be fully resolved until such work is carried out, but in microbiology, in some botanical groups, and also

in entomology (Michener's own field), numerical taxonomy has been used precisely for these purposes, i.e., as an exploratory tool in several groups of organisms that previously had not been studied. The inevitable development of automatic character recording devices may also outdate this particular argument.

It is significant that there have been rather few criticisms of numerical taxonomy applied to bacteria. The main ones are by Adams (1964) and Leifson (1966). Adams criticizes equal weighting of characters but proposes no practicable alternative. Leifson's main objection is that certain classes of characters, particularly morphological ones, were overemphasized in some published studies, an assertion that appears extremely dubious. The great majority of studies have given acceptable taxonomic results. Why then do some critics maintain that numerical taxonomy does not work in higher organisms? One might anticipate that bacteria were particularly difficult organisms to classify. The reasons given by Blackwelder (1967b) that the characters are those of bacterial populations and are inherently quantitative seem to have little force: the characters of single cells (if available) would doubtless be much the same, while the nature of bacterial characters cannot be said to be inherently either qualitative or quantitative (the qualitative characters are due in the main to dichotomizing quantitative characters at some level convenient in chemical assay). It is true that homology in bacteria is mostly based on external criteria, and the implication of this for other organisms has not yet been explored. It is also true that most characters are not morphological but biochemical or physiological, yet it would be hard to argue cogently that their significance was very different, particularly since no very obvious peculiarity in taxonomic pattern of chemotaxonomic characters has been noted in higher organisms. If dichotomizing characters avoids some problems (like that of the general size factor), it may be profitable to use this method in higher organisms. But it would seem to us more likely that the main reason for the general acceptability of numerical phenetics in bacteria is that the user is expecting only phenetic groups and is satisfied when he obtains them (see Sneath, 1971b).

Criticisms of Empirical Findings

Typical of such publications is a critique of the paper by El-Gazzar et al. (1968) in Burt et al. (1970), in which I. C. Hedge points out that the classification comes out in an unexpected way. Yet no reasons other than those of preconceived notions regarding the genus are given: "the underlying reason for this is not hard to find. The characters used were unsatisfactory for the genus... In any future study of *Salvia* ways would have to be found to codify the features of the corolla form in which much of the diversity of the genus is expressed..." These statements summarize the preconceived notion that "experienced taxonomists" know what characters are useful for describing species and genera, and that any studies are obviously wrong which do not use these characters or which use these characters

intermixed with others that these experts have not found to be equally useful. Numerical taxonomy and its philosophy have attempted to escape this kind of dogmatism. The critique of a study of *Ericales* by Watson, Williams, and Lance (1967) by P. F. Stevens in Burt et al. (1970) is of greater interest because some of the comments are clearer and more relevant. The criticism of the choice of characters remains philosophically unacceptable to us, but Stevens engages in a cogent discussion of the delimitation of character states and of errors of observation (see also Stevens, 1970). As we have discussed elsewhere in this book (Sections 5.10 and 5.13) errors in such work will necessarily result in classificatory errors of possibly predictable magnitude.

The criticisms also raise another point: what are the reasons for the failure of the numerical taxonomic method with certain OTU's when with others the phenetic groups were quite acceptable? What characters would have to be recoded, differentially weighted, and so on, to give results that would be acceptable with all the OTU's? If we could find this out we would be better able to avoid such errors, or the alterations in the data would have to be so extreme as to cast doubt on our preconceived ideas. Indeed, Brown (1965) illustrates this last point in asking whether the impressive list of common characters of Strepsiptera and rhipiphorid beetles is due to convergence or close common ancestry; the problem is seen to turn on whether it is more probable that there has been a gain rather than a loss in the number of tarsal segments, a point that is evidently very hard to decide. Similar comments are made by Blackith and Blackith (1968) who say, in speaking of the results of a numerical taxonomic study, "there do not seem to be any grounds for accepting only those parts of the picture which are easily assimilated and rejecting those parts which are less so. To select features in that way would be to condemn numerical taxonomy for being wrong when it disagreed with accepted ideas and superfluous whenever it agreed."

An important point to be stressed here is that the explicit procedures of numerical taxonomy make it possible to engage in criticism of this sort, which is not generally possible in orthodox taxonomy. Many if not all of the criticisms of methodology and results of taxometric studies apply to conventional classifications as well but cannot be stated as clearly in that branch of taxonomy. In other words, the very same criticisms could be directed at the treatment of the problem by conventional taxonomy, except that the problem did not become obvious until it was treated by numerical taxonomy. As just one of many possible examples we cite Steyskal (1968), who feels that many more than the usual number of characters may be required for numerical taxonomy in certain insect groups in which taxa can only be recognized on details of genitalia. Because the nonspecificity hypothesis does not hold in its entirety, it is possible that inadequate sampling of characters may distort phenetic relationships among such groups of insects, but this danger would affect conventional and numerical studies equally.

L. A. S. Johnson (1968) states that the effort expended in numerical phenetics is not worthwhile in view of the published results. We believe he underrates the significance of the findings that have been made and that we have illustrated in our present book. There is hardly a numerical taxonomic study that has not resulted in some new insight and understanding and there are many that have been of considerable heuristic value.

We share with L. A. S. Johnson (1968) his desire for a heuristic element in taxonomy and we shall show in Section 10.3 that numerical taxonomy can help pose some important biological questions and provide some answers, however automated the procedure of character gathering, defining, and classifying eventually becomes. We should clarify our view and state that we do not thereby imply that *only* numerical taxonomy has these heuristic properties. There are numerous ways of looking at biological systematics and using it to generate hypotheses about a variety of biological problems. Many of these can and will be quantified, but not all of them can be or need to be.

10.2 SHORTCOMINGS OF NUMERICAL TAXONOMY

Throughout the text we have pointed out the various problems encountered when one attempts to make taxonomic procedures operational. As in all sciences, the solution of such problems is a never-ending procedure. Having satisfactorily overcome the obstacles of one aspect of the work, new difficulties emerge at another level. It may be useful to bring together in one place the various problems discussed in the preceding sections that still await solution.

Perhaps the foremost challenge is the development of a generally acceptable system for coding and scaling characters. The problems involved have been outlined in Section 4.8. Possibly the best hope lies in analyzing in some way the information content provided by various characters, followed by their reduction into binary unit characters. The difficulties posed by logical dependence and biological correlation are manifold. Until such time as a general system for describing characters is available, workers must be content with the present methodology, which in spite of the lack of elegance of the theory is fortunately robust in its results. Uniform coding of characters will of its own accord lead to a type of automatic weighting. Characters conveying more information will contribute more toward the estimate of resemblance between OTU's.

Clearly more work is needed on defining the character sets that can and should be used in taxometric work, and more precise and profound analyses of the nature of phenetic similarity are required. We have already departed from the concept of a parametric overall similarity (Section 3.8). The difficulties inherent in such a concept have been enunciated very cogently by L. A. S. Johnson (1968). He

furnishes the following list of subjective decisions that still must be made when a measure of similarity is adopted :

- (i) the set of objects considered usefully comparable ;
- (ii) the domain of attributes which we consider relevant to our interest in the objects ;
- (iii) the "fineness" with which we analyse the features into elementary attributes (\equiv states) ;
- (iv) the establishment of equivalences or homologies between parts of the objects under comparison ; and the consequent grouping of the elementary attributes into two- or multi-state sets, thus specifying what we usually term "the attributes" or "the characters" ("multi-state" here includes "continuously-varying") ;
- (v) the method and intensity of sampling of the objects ;
- (vi) the method and intensity of sampling of the acceptable sets of "relevant" attributes ;
- (vii) the quantitative or qualitative measures to be used in expressing the "states" of each attribute (involving an often arbitrary assignment of working commensurability) ;
- (viii) the measure of similarity (or distance) to be adopted.

Similar problems arise with apparently attractive alternatives in the study of proteins or DNA sequences. Much work lies ahead in these areas. It should be added that most of these problems face the traditional taxonomist as well.

Tests of significance of similarity coefficients (discussed in Section 4.10) and of classifications (Section 5.10) will undoubtedly progress considerably in the next few years and will, as a consequence, bring about profound changes in present day procedures.

The discussion of taxonomic structure in Chapter 5, especially optimality criteria (Section 5.10), has pointed out the difficulties that still beset any definition of an optimal classification. Such a definition will comprise decisions on the kind of structure needed by taxonomists, desirable mathematical properties of optimality criteria, and suitable algorithms for computer handling.

In this connection we may again cite L. A. S. Johnson (1968), who provides a large and impressive list of problems to be solved by numerical taxonomists. He furnishes this list as an example of the tasks that should occupy numerical taxonomists, with the implication that such activities are aside from and possibly have no relevance to the problems of taxonomy. Yet we feel that the cogent questions posed by him are the very questions to which any taxonomist needs answers if he is to understand the natural world that he wishes to classify.

What we consider the primary shortcomings of numerical taxonomy have been listed above, but we might round out our account with some problems of secondary importance. As should be evident from the extended discussion in Section 3.4, we still lack a consistent system of operational homology. It appears that phenetic analyses can be carried out in spite of this drawback, but nevertheless it is philosophically embarrassing to rest character coding on such shaky foundations. The importance of correctly defining homologies becomes especially relevant in cladistic work and in evolutionary studies based on protein sequences, and recent

work has emphasized this problem. We still know little of the causes of incongruence; methods for partitioning the variation into components of biological interest, and accounting for sampling error are thus much needed. Some patterns of variation have not been looked for by numerical pheneticists, such as that in which two taxa are sharply separated on a few characters only, so that they form, in phenetic space, two flat plates separated by a small gap. This might occur with pairs of sibling species, and the existence of two such taxa would not be readily detected by the usual methods of finding taxonomic structure. In Section 6.3 we have stressed the importance of correct assumptions about evolutionary processes in order to make cladistic inferences for numerical cladistics and numerical phyletics. In order to construct explicitly analytical algorithms we must have some understanding of how to reconstruct ancestral character states and the degree of confidence we can have in such work. We are in a period of rapid change in our techniques in this field at the moment and readers are referred to Kluge and Farris (1973) and to the numerous articles appearing on this subject in the technical literature.

10.3 HEURISTIC ASPECTS OF NUMERICAL TAXONOMY

The term "heuristic" is defined as "revealing new knowledge." Specific examples have been given in other chapters of the new information that can be gained from numerical taxonomy, so we are more concerned here to point out the kinds of new knowledge that can be thus obtained, rather than details. The discussion includes some aspects of related multivariate techniques, because new knowledge is often obtained from a combination of these with the more narrowly taxometric methods.

Several authors have noted the potential of numerical taxonomy for forming new hypotheses. Thus Silvestri and Hill (1964) point out that the discovery of well-separated phenons implies that in nature intermediate forms will be rare, which can be tested by further work. They foresee that taxonomy will cease to be a purely descriptive science but will also be used for formulating hypotheses that can be tested experimentally. Similar views are expressed by Williams and Dale (1965), Williams and Lance (1965), and Goodall (1966e). This has, of course, always been true of taxonomic work to a certain extent, but its power will be greatly increased by quantitative techniques. Such work may have important economic consequences, especially in such fields as agriculture and public health. For example, the association between streptomycete taxa and the antibiotics they produce is of importance to the pharmaceutical industry (e.g., Gilardi et al., 1960; Silvestri et al., 1962). The observation of clusters of OTU's or of characters that are empirically correlated may lead to a search for a causal explanation. Thus floral character correlations may be causally associated with pollination mechanisms. Blackwelder (1967a) and Mayr (1969a) give numerous other examples of the practical importance of taxonomy.

We may first consider what general taxonomic information may be obtained from numerical studies. The most obvious is that the numerical taxonomy confirms the broad outlines of existing orthodox taxonomy. Although this may come as no surprise, and may sometimes be categorized as "self-evident information," such confirmation should not be despised. It increases the general credibility of the taxonomy of the group and of the numerical techniques appropriate for such study. Numerous examples could be cited, e.g., Michener and Sokal (1957) on bees; Sims (1966) on earthworms; Wilkinson (1967, 1968) on lepidoptera; Ackermann (1967) on birds; Watson, Williams, and Lance (1966), Rhodes et al. (1968), Stearn (1968, 1969), Ivimey-Cook (1969a,b), Prance, Rogers, and White (1969), and Rowley (1969) on flowering plants; and Seyfried (1968) and Stevens (1969) on bacteria. More to the point is that in many cases there have previously been several conflicting orthodox taxonomies, and the numerical analyses have often provided strong support for one of these. The studies just mentioned contain several such examples.

It is very common to find minor points of dispute in the previous taxonomy of a group, affecting only a few of the smaller taxa. Numerical studies have been extremely useful in giving more convincing evidence of the phenetic relationships and the better placement of such taxa (for examples see Hudson et al., 1966; Thornton and Wong, 1967; Rowley, 1967; McNeill, Parker, and Heywood, 1969a). Numerical taxonomy occasionally suggests quite unsuspected relationships (e.g., Barnes and Goldberg, 1968, on anaerobic bacteria). Again, some accepted groupings have been shown to be of dubious validity; for example, Stephenson, Williams, and Lance (1968) were unable to substantiate the existence of some "twin species" that had been proposed in orthodox classifications.

Numerical studies have been particularly informative in groups whose current taxonomy is very confused or is refractory to orthodox methods. Most often this has suggested major revisions of the taxonomy. Such confusion is most widespread in microbiology, where many examples can be found (reviewed in *Principles of Numerical Taxonomy* and in Sneath, 1964a). In higher organisms we know of fewer examples, but the studies of Ducker, Williams, and Lance (1965), Wilkinson (1970a), El-Gazzar and Watson (1970a,b), and El-Gazzar et al. (1968) may be cited. In the last study, on the genus *Salvia*, of the existing subdivisions only one of five subgenera and only one of twelve sections appeared valid, though this study has been criticized by Burt, Hedge, and Stevens (1970); the numerical results yielded new and more convincing phenons in close accord with geographic distribution and with several additional characters. An even more drastic reassessment of current views is posed by the study of Young and Watson (1970) on the families of dicotyledons. An allied problem is to extract some sense out of data of very low quality, and Parker-Rhodes and Jackson (1969) give an example. In bacteriology numerical taxonomy has been of great value in the initial grouping together of individual strains to yield taxa of low rank conventionally regarded as species, and most of the work on bacteria deals in part with this problem. One common result

has been to simplify greatly the taxonomy of bacterial taxa by reducing many dubiously named species to synonymy, but sometimes numerical methods have revealed subgroups considered to be valid new species within apparently homogeneous groups of strains. There have been a few studies in higher organisms where numerical taxonomy has been used to find new taxa without first attempting an orthodox taxonomy (e.g., Imaizumi, 1967, on a new genus of cat).

Although numerical taxonomy has occasionally increased the number of rank categories that may be profitably recognized, we do not feel that it has resulted in any marked tendency toward either "lumping" or "splitting" when all kinds of organisms are considered. It has however greatly improved our knowledge of the finer points of relationship.

The results of numerical taxonomy have often been confirmed by subsequent work. This is particularly true in bacteriology (reviewed by Jones and Sneath, 1970) and some such confirmatory work on higher organisms has been mentioned in Section 5.12. Additional examples may be found in the studies of Hudson et al. (1966), Drury and Randal (1969), and McNeill et al. (1969a) to mention only a few. But although confirmatory evidence has not always been presented, it is noteworthy that cases where clear and apparently acceptable numerical results have been contradicted by later work seem to be uncommon. We find it difficult to cite good examples of this. In most cases the discrepancies appear to be due to inadequate techniques, or methods inappropriate to the desired taxonomic aims. Some of these are discussed in Sections 10.1 and 10.2. We have noted, however, in common with others (e.g., Moss, 1968b; Rhodes et al., 1968; Huber, 1968; Johnson and Holm, 1968), that many minor discrepancies may well be due to size factors (see Section 4.11), which therefore deserve careful attention.

More specific information that may be obtained from numerical studies includes the prediction of states of characters that were not included in the original survey. A good example is given by Williams (1967a). In the study of Ericales by Watson, Williams, and Lance (1967) the genus *Epigaea* (always something of a puzzle) was placed in the Phyllodoceae. But members of this tribe have an unusual feature, the presence of viscid threads among the pollen grains, which was not recorded for *Epigaea*. On further examination of *Epigaea* the threads were found. This is, of course, a special case of the general congruence principle that good phenons are likely to be relatively homogeneous for states of new characters, of which Drury and Randal (1969) give an example. Numerical studies may also reveal the insufficiency of the available characters for adequate taxonomic work, e.g., as revealed by large errors of sampling or measurement or because unsatisfactory groupings emerge (see Sections 3.8, 4.10, 6.4).

Congruence and incongruence between phenetic relationships based on different life forms is another field where quantitative techniques are indispensable. Such work could give new knowledge of the probable identity of larval stages and adult stages (e.g., in parasites) or of the sexual and asexual forms of fungi (e.g., Leth-Bak

and Stenderup, 1969). The grouping of fossil fragments into associations representing different species has been noted in Section 6.5. More important, the measurement of incongruence requires new hypotheses to explain it. Thus larval-adult incongruence will some day illuminate aspects of the evolution of these stages, the selective pressures and adaptations involved in the habitats of larvae and adults, etc. (Rohlf, 1963a). Similarly incongruence between floral and vegetative characters (Crovello, 1969; Johnson and Holm, 1968) must have biological meaning. Pernes, Combes, and René-Chaume (1970) proposed a hypothesis on inheritance mechanisms in millet as a result of observing incongruence. Incongruence often seems to affect mainly a few of the OTU's (Moss, 1967; Ehrlich and Ehrlich, 1967; Johnson and Holm, 1968), a finding which is still poorly understood.

Other details obtainable from numerical studies include patterns of phenetic variation. Thus phenons may be found to be hyperspherical or elongated, they may show trends or clines in ordination studies, or chains of intermediate forms may be revealed (see, in particular, Sections 5.14 and 7.5). It is also to be noted that satisfactory schemes of discrimination may require meticulous descriptions of phenons in numerical form; this can be a critical step in difficult situations (e.g., Lapage et al., 1970). The use of phenetics in nomenclature, for choosing types and limits of taxa, has been discussed in Chapter 9. New knowledge on morphogenesis, growth, functional analysis, and the interaction of genotype and phenotype can also be obtained from phenetic studies (Blackith, Davies, and Moy, 1963; Sneath, 1967a, 1968a; Soulé, 1967b; Oxnard and Neely, 1969). Borgelt (1968) discovered by numerical methods a new form in the solitary stage of the life cycle of a salp that corresponded to a known form in the aggregate stage.

In the field of evolution the growth of numerical cladistics has opened up a large new area. Numerical phenetics and cladistics together can allow the measurement of rates and patterns of evolution, selective pressures, the study of convergence, parallelism, and reticulate and mosaic evolution. These points are fully discussed in Chapter 6, and here we need merely add that studies of this kind can only be accomplished by numerical methods. Broader information includes generalizations about the way evolution operates, concordance between geological time and divergence (both morphological and biochemical, as in serology and protein sequences), and the logic of cladistic inference.

Phenetic studies of various kinds have been valuable in genetics. Thus Whitehouse (1971) has used multivariate methods to predict the parents most likely to yield desired hybrid crop plants. Rhodes, Carmer, and Courter (1969) have employed numerical taxonomy to select the widest range of variation for programs of plant breeding and Rhodes and Carmer (1966) discovered errors in pedigree records as a result of unexpected phenetic relationships. Incongruence between phenetics and ability to hybridize may be an interesting field (Heiser, Soria, and Burton, 1965). Numerical phenetics can give new information about the effect of environment on phenotype (an unexplored field); Williams and Lance (1969)

mention the use of cluster analysis in elucidating a complex problem of this kind.

Geographic and ecological variation has been discussed in Chapter 7, but we may summarize this by saying that phenetics gives abundant information on the pattern of geographical variation, the reality (or otherwise) of subspecies and races, and the occurrence of temporal phenetic changes (Mason, Ehrlich, and Emmel, 1968). Numerical taxonomy may provide information of especial value in epidemiology. Thus Ibrahim and Threlfall (1966a) found a very close phenetic resemblance between isolates of a fungus from oats and from rye grass, and comment that this appears to solve the problem of the source of secondary infection of oats. Talbot and Sneath (1960) noted a correspondence between phenetics and pathogenicity of the hemorrhagic septicemia bacillus: all the strains from cats and from internal lesions in humans were extremely similar but differed from most dog strains; this suggests that human infections with this bacillus are usually contracted from cats. Roberts (1968) found certain subclusters of *Corynebacterium pyogenes* associated with different hosts, and more weakly with the organ infected and with geography. Jewsbury (1968) discovered complex relationships between host, geography, and possibly also snail vectors in a study of schistosomes.

Applications of numerical methods outside systematics are discussed at length in Chapter 11. We mention here only a few examples that illustrate the kind of new knowledge that may be obtained. Most work has been done in ecology. This may reveal quite unsuspected information. A floristic study may uncover a discontinuity due to a change in soil, a difference in agriculture, or the effects of burning of vegetation. Detailed ecological study of forests is giving an entirely new view of a complex habitat. These developments are reviewed by Williams (1967b), Williams and Lance (1969), and Webb et al., (1967a,b). Major biogeographical patterns may be well shown by numerical methods, and invasion routes may be indicated (e.g., Holloway and Jardine, 1968; Berry, 1969). Climate and the geological history may be reflected in ecological resemblances (Proctor, 1967; Sneath, 1967c). Other fields where phenetic studies are providing new knowledge are also discussed in Chapter 11 (e.g., stratigraphy).

A major contribution to evolutionary and ecological theory has been the study of biological diversity initiated by Margalef (1958) and widely applied by numerous authors in recent years (e.g., MacArthur and MacArthur, 1961; Pianka, 1967; Johnson, 1970). Diversity indices have so far been based on numbers and frequencies of species—the traditional species of conventional taxonomy. The tacit assumption of these formulations is that each species is ecologically equivalent. Yet to weight such species occupying the ecological space in these faunas and floras equally in a diversity formula is clearly unrealistic. One looks forward therefore to a new field of research in which the study of ecological diversity will be combined with numerical taxonomy and in which diversity measures will include indices of the hypervolumes and the density of the various species concerned. Initial attempts at numerical taxonomic definitions of niches have been made by Hutchinson (1968),

Fujii (1969), and Maguire (1967). Hurlbutt (1968) made an attempt to illustrate Gause's principle by means of phenetic differences between species coexisting in the same habitat. Other work in this field is in progress (e.g., Findley, 1973).

Finally, numerical taxonomy has a very important part to play in scientific methodology. Indeed this alone would justify its study as a scientific discipline (as admitted by its critics). It has given deeper insight into the nature of the taxonomic process and the nature of taxonomic judgment. Concepts such as size and shape have only become susceptible to effective study as a result of numerical methods. Many questions, often awkward for taxonomists, are raised: what is a taxon?; what is a character?; what is resemblance? Whether or not the answers of numerical taxonomy have been useful, its questions certainly have been. It will contribute toward the psychology of perception. It seems from the little present work (see Hodson, Sneath, and Doran, 1966; Sneath, 1968b) that trends, clines, or progressions are much more difficult to detect by eye than are discrete clusters when variation is complex. There are some suggestions (from examples in Sneath, 1964b) that the eye is often misled to overvalue an attribute that is large or affects a large area (such as prominent coloration), while position, stance, and orientation are less often misinterpreted. Several studies currently underway (Moss, 1971; and the TAXOCRIT experiment by Sokal and several associates) are beginning to answer questions about what taxonomists actually see in making their judgments. An interesting essay on the methodology of classification and perception is that of Ornstein (1965). The nature and value of taxonomic experience has been investigated in a first study by Sokal and Rohlf (1970). The study of automatic scanning methods (Sokal and Rohlf, 1966) is a logical extension of work on perception. Numerical taxonomy contributes also to the study of homology, where work is being actively pursued both on gross morphology and protein sequences.

In conclusion we must emphasize that though many of the kinds of information discussed in this section may seem to be self-evident from the crude data, this is not by any means so in most instances. We may cite here a comment from an anonymous contributor to the discussion on the paper of Willmott and Grimshaw (1969), who states: "One simple way of checking on the production of 'self-evident' results, which I have myself used with ecologists in the application of association analysis is to ask the ecologist to write down his interpretation and place his description in a sealed envelope before starting the analysis. The interesting result is that in every case the analysis gives the information which a trained ecologist will derive from his knowledge of factors external to the data, and that in most cases it will find something which that ecologist, despite his extra knowledge, has missed." Our personal experience in a variety of other fields is virtually identical.