

Taxonomic Evidence

We now proceed to a detailed discussion of the data necessary for obtaining estimates of resemblance between taxa. This leads us first into a consideration of the fundamental taxonomic units employed in numerical taxonomy, then to a discussion of the nature of taxonomic characters, questions of homology, and problems in character sampling and evaluation, such as choice and weighting of characters.

3.1 OPERATIONAL TAXONOMIC UNITS

In order to restrict one's study to a manageable taxonomic group, one must first make a preliminary selection of specimens. No sharp distinction is made between the selection of specimens, discussed in some detail here, and the selection of characters treated in Section 3.6, since these generally proceed *pari passu*. Inasmuch as a taxonomic group is selected by "classification from above," selection is therefore necessarily based on rather few characters.

A point of some importance is to guard against the exclusion of pertinent material because it does not strictly fulfill all the criteria for the working definition of the taxon to be studied. The danger is that aberrant forms or descriptions of aberrant taxa in the literature may be excluded because they do not possess some character of the taxon that the systematist considers essential or diagnostic. The

inclusion of a small amount of possibly atypical or unsatisfactory material, which may have to be excluded in the final analysis, is a worthwhile insurance against an unrepresentative study. Very similar considerations will apply to material that has been extracted from the literature, though the danger of mistakes is naturally greater.

There is no need to belabor the point, now very well understood, of choosing adequate numbers of specimens for establishing taxa. These could be individuals, for establishing species, or species, for erecting higher taxa. To some degree the exemplar method proposed in the next chapter (Section 4.13) is in contradiction to this admonition, since by that method only single or a few representatives of given taxa are used in the studies. However, it is clear that such exemplars are only reference points and do not indicate the limits of the taxa which they represent. The special problems attendant on material of different ages, different stages of the life cycle, and fragmentary material are discussed in Sections 4.9 and 4.12.

What taxonomic units can be classified by numerical taxonomy? The logical fundamental unit in a large majority of instances is the individual organism. Hennig (1966, p. 6) points out that the fundamental element of biological systematics is an individual at a particular point of time. He calls this unit the "character-bearing semaphoront." For purposes of systematics one must study the multiplicity of characteristics of a sequence of semaphoronts, yielding a description of the phenetics of a life cycle, the holomorph in Hennig's terminology. The individual (or the individual holomorph) is usually an unambiguous entity.

In some studies individuals are the basic units employed. Such studies would throw light on resemblances among intraspecific variants, but would not be likely to offer much scope for comparisons at the subgeneric, generic, and higher levels. Also the comparison of numerous individuals would lead to similarity matrices of excessive size and attendant difficulties of computer processing. Consequently, at higher taxonomic levels representative specimens (the exemplar method, Section 4.13) or averages of several individuals are employed for a taxon. In many studies the unit taxon will therefore be an individual (or an average) standing for a race, a species, a genus or even a higher ranked taxon. The most common unit in classificatory studies will be the species (strictly speaking, the taxonomic unit with a binomial name which is believed to correspond to one or other of the biological units which are given the name of species). Thus we shall employ taxonomic units of different categorical ranks as the entities to be grouped into more inclusive aggregates during classification. We cannot therefore speak of fundamental taxonomic units (these would mostly be individuals), but shall refer to *operational taxonomic units (OTU's)*, which are the lowest ranking taxa employed in a given study. From study to study OTU's can therefore differ in rank; for example, they may be individuals, exemplars of genera, or averages representing species.

It is important to avoid prejudice in choosing the OTU's and, if need be, to explore, by preliminary analyses, the phenetic relations of the specimens that are

to form them. In both plants and animals there may be a choice between stages in the life cycle, or there may be a choice between life forms, e.g., castes in social insects. The special problems raised by these possibilities will be discussed in Section 3.6.

Should numerical taxonomy rely on the validity of prior classification for its choice of OTU's? To be totally consistent and rigorous, a taxonomic study at any level should be based on individuals. But since not all individuals of every included population can be studied and analyzed, sampling of individuals is necessary. Williams and Lance (1965) have discussed at some length the general problem of choosing a sample of individuals, and they point out that while it is desirable to have a representative sample (so that the study can be probabilistic) this is not always possible, and nonprobabilistic studies are perfectly valid within their limitations. Also, if we wish to reexamine the relationships of numerous genera in a family we cannot reassert the validity of every genus from a study of its species (and the validity of the species from a study of their individuals). If validation at every level were attempted, preliminary studies would take so much time that the originally intended classification would never be accomplished. Operationally, we have to take on faith the validity of OTU's above the rank of individuals, to "lift ourselves by our bootstraps" into a position from which it is possible to carry out classification at the intended level.

Of the supraindividual categories, the species is probably the most frequently encountered. Although this is claimed to be a nonarbitrary category by adherents of the New Systematics (Mayr, 1963, 1969a,b), it can be shown that the definition of the so-called biological species is nonoperational and that effectively all species are so considered on phenetic grounds (see Section 7.1, and Sokal and Crovello, 1970). The employment of species as OTU's in a phenetic classification is therefore appropriate.

Problems may arise if a taxon used as an OTU proves to be variable for one or more characters. This brings up the question of whether, as OTU's, we can use higher taxa (such as genera, families, and orders), in which the majority of characters within a taxon will, of course, vary. Such taxa can be used, in principle, for the reasons discussed in Section 4.13. A second equally serious problem is the low degree of relevance of most lists of characters (see Section 4.12).

One practical consideration is what to do when the number of OTU's is unmanageably large. A solution that is frequently successful (Sneath, 1964a) is the following: run a random sample of the OTU's and from each well-defined cluster pick three OTU's as reference points for that cluster. Run a second sample including these reference OTU's, and repeat until all the OTU's have been run. Many OTU's will be seen to belong to clusters previously recognized. All the remaining "solitary" OTU's should preferably be rerun with the reference OTU's to detect smaller clusters. Experience with the exemplar method (Section 4.13) gives confidence in this procedure. Three reference OTU's per cluster also provide an internal check on the study, in that they should always cluster together closely. It may be noted

that although the exact composition of the set of OTU's has some effect on the phenetic relationships that are observed, numerical taxonomies are quite robust to small changes in the composition, as shown by Crovello (1968f). This is taken up further in several sections of Chapters 4 and 5.

3.2 DEFINITION OF TAXONOMIC CHARACTERS

Procedure in taxonomy, orthodox or numerical, is based on taxonomic characters. The term character has been employed in at least two distinct senses by systematists. Its commonest usage is as a distinguishing feature of taxa—a characteristic (or feature) of one kind of organism that will distinguish it from another kind. Thus, serrated leaves may distinguish one species of plant from another and hence are called a character; similarly, punctate elytra may be used to differentiate between two species of beetles, or resistance to phenol may separate two strains of bacteria. This appears to be the sense in which Mayr (1969a, p. 413) defines a taxonomic character, as “any attribute of a member of a taxon by which it differs or may differ from a member of a different taxon.” Unfortunately, such definitions point out again the dilemma of conventional taxonomic procedure: characters are restricted to differences between members of taxa, but the taxa cannot be recognized without the characters themselves being first known.

Another frequent meaning of the term character, which has been espoused by numerical taxonomists as being the more useful one in their work, is that a character is a property or “feature which varies from one kind of organism to another” (Michener and Sokal, 1957) or “anything that can be considered as a variable independent of any other thing considered at the same time” (Cain and Harrison, 1958; we assume the independence referred to is logical rather than functional or mathematical). Thus, referring to the previous examples, the nature of the margins of the leaves becomes the character, while entire, serrated, undulating, or any other types of margin become different *states* of the character (Michener and Sokal, 1957). The word “state” may imply qualitative rather than quantitative subdivision, but in the absence of a more suitable term we employ it to cover both. In the sense in which these terms are used in this book, the condition with respect to punctuation of the elytral surface and the property of phenol resistance would be characters, while smooth or punctate and resistant or susceptible, respectively, would be character states.

A note of caution: Blackwelder (1967a) would refer to our “states” as characters and Colless (1967a) prefers to call them attributes, using character in the same sense as we do. Readers of N. Jardine's numerous important contributions to taxonomic theory should be aware that these same terms are used in yet a different sense by him (see Jardine, 1969b). The descriptive terms applied to individual organisms are called *attribute states*, while sets of such descriptive terms are called *attributes*. The term *character state* is used by Jardine to designate the probability distributions

over the states of an attribute, while sets of such distributions are called *characters*. His terminology is unfortunate inasmuch as the meanings of the terms character and character state are by now well established in numerical taxonomy; furthermore there is nothing in these terms to connote to the casual reader the idea of a probability distribution.

No objection should be raised to defining a character as a feature which varies from one organism to another. However, if we say that it varies between kinds of organisms (or species) then we are ourselves in the same sort of dilemma as a systematist attempting to apply the definition of character in the first paragraph of this section—that is, defining characters on the basis of predefined taxa. Thus we would first have to define our species before we could define the characters. To be extremely critical, therefore, we would have to define characters entirely on the basis of the differences between individuals. Specific characters are, of course, nothing but summaries or abstractions of the characteristics of a large number of individuals.

The general definition of characters established above cannot, however, give much aid to the practicing taxonomist in the process of recognizing and describing individual characters. We shall go into this problem in the next section, from both theoretical and practical points of view.

3.3 UNIT CHARACTERS

Theoretical Considerations. Those embarking on work in numerical taxonomy may be puzzled by the task of recognizing the basic units of information for the study, what we call the *unit characters*. In trying to define these characters we must first ask ourselves what properties we wish to recognize. Do we wish to recognize genes, or a unit element in selection, or a logical construct? If, as is now clear, genes are themselves complex entities, shall we subdivide them? And if so, to what extent?

Most definitions of a unit character have been too restricted; defining a unit character in terms of morphology, chemistry, genetics, or evolution does not allow the broad treatment needed for a general theory of systematics. For this we need to define unit characters in terms of information theory, for in every instance it is information that the characters convey to the taxonomist; this idea is closely linked to the concept of natural taxa as groups with high content of information (see Section 2.2). Sneath (1957b), has made an attempt to introduce the concept of information, and a unit character (there called a “feature”) was defined as an attribute possessed by an organism about which one statement can be made, thus yielding a single piece of information. These attributes are formally logical constructs, since they will change if the technique of observation changes; the definition is therefore an operational one. If a character can vary continuously, such as with the length of an organ, the character of length can be broken down into as many steps as the observational method will allow with good reliability. Either each step

is counted as a feature, or at least the minimum number of features necessary to account for the existing variation is postulated (see Section 4.8). But as is also pointed out below, these steps will not be independent, since they are not mutually exclusive.

This approach may be carried to its logical conclusion, where each unit character or feature represents an alternative that can be answered as “Yes” or “No,” “Possessed” or “Not Possessed”; the information content can then be measured as “bits,” as is usual in information theory.

Clearly even the most simple organisms contain a great many bits of information. We may plausibly interpret this information in terms of modern genetics. The genome consists of a series of nucleotides, which are paired one-to-one in a double helix of deoxyribonucleic acid (DNA). The genetic information resides in the sequence of the different nucleotides and may be thought of as a code message written in an alphabet of four letters, each letter representing one of the four alternative nucleotides—those containing thymine, cytosine, adenine, or guanine. The genetic code message is translated into other codes determining the amino acid sequence of proteins and the structure of other molecules in the cell, and these in turn determine the physiological and morphological properties of the organism. Although evidence is still scanty, there are two lines that confirm the close correspondence of the fine genetic structure with orthodox and numerical taxonomy. One is the data from protein sequence studies and the other is the congruence between phenetic data and DNA pairing (Section 3.6). There are, however, some difficulties in relating the fine structure of the genome, even if it were completely known, to taxonomy. Thus, if there has been a gene reduplication, should this be scored as many independent characters, or as one change, the reduplication itself? This is clearly analogous to estimating the common information (in the conventional sense of the term) in two pieces of written text, such as two editions of a book, where insertions, repetitions, and changes of order may occur.

We may tentatively identify our taxonomic bits with the genetic code and in *Principles of Numerical Taxonomy* we noted that the number of bits in the genome ranges from around 10^4 for some viruses to around 10^{11} for many higher animals and plants, assuming that the bulk of the DNA is functional. The number of functional genes is of course much smaller, since each gene is made up of many bits, probably about 1,000 as a rule. These figures are only speculative, and do not include any nonchromosomal genetic information or environmental effects, but it is clear that the content of information (in the information-theoretical sense) is much smaller in microorganisms than in higher organisms and is related to histological complexity (Sneath, 1964d). The scale of the potential store of information in a nucleus may be judged from the estimate that the Library of Congress contains between 10^{13} and 10^{14} bits (see Good, 1958). It is interesting to note that Elsasser (1958, pp. 100–104) estimated from morphological considerations that the information in man was at least 10^7 bits. It should be noted that the bits are a measure of

the potential information content, not of the number of alternative permutations of the information. The latter is 2^x when the number of bits is x .

Working Definition. Except in those few and simple organisms whose fine chemical structure is gradually being unraveled, the above considerations are premature, so a working definition, which the practicing taxonomist can use, is needed. We may define a unit character as *a taxonomic character of two or more states, which within the study at hand cannot be subdivided logically, except for subdivision brought about by the method of coding.*

Since we cannot in most cases make genetic inferences from phenetic studies of characters, we shall generally have to use phenotypic characters as our basic information, defining these as narrowly as possible. Our failure to make *logical* subdivisions may rest on ignorance of the finer structure or the causation of a character. Thus, presence or absence of a bristle in an insect may be a unit character, if we know nothing of its finer structure and have no way of subdividing it, or do not care to do so. Even if the general morphogenesis of the structure is known from a representative form, unless morphogenetic differences can be established that can serve as taxonomic characters within the group studied, the presence or absence of the bristle remains the unit character. To consider another character, the same insect may possess or lack DDT-dehydrogenase, which character we are again unable to subdivide further in view of our present knowledge. Thus, the organizational levels of unit characters may differ considerably from character to character and with advances in our knowledge. The ruling idea is that each character state should contribute one new item of information.

We anticipate in the near future that automated methods will be introduced for extracting information directly from specimens and converting it into unit characters. Such automation is now being introduced in some fields of biochemistry and microbiology, but its application to morphology may prove more difficult. A step in this direction has been taken by Sokal and Rohlf (1966) and Rohlf and Sokal (1967) who showed that remarkably good numerical taxonomic results could be obtained by overlaying biological images with a perforated mask. If a line in an image showed through a given hole this was scored as a character with value 1, and if not, with value 0. The two sets of characters were then used in a numerical study. As the method stands at present, it must be assumed that two images have been appropriately oriented and scaled for size. To generalize the technique these factors will require attention, but they do not appear insuperable problems, although some points regarding operational homology require further development (see Section 3.4). Recently Meltzer, Searle, and Brown (1967) described a method of defining large and small scale detail of shapes by means of rectangular reference patterns, which may offer another approach toward automatic scanning. There are also potential developments based on holography (Gabor, 1965).

When corresponding (operationally homologous) points are used for the method of matching images (Sneath, 1967a) the definition of unit characters as such is circumvented, and the positions of the points effectively become the unit characters. The dissimilarity, d_n , between two images is then obtained as a form of average distance between the pairs of corresponding points (one from each image) at the orientation of the images that gives the best possible match.

3.4 HOMOLOGY

A Critique of Traditional Definitions

The concept of homology is central to any taxonomy. All nonarbitrary classifications are based upon comparisons of sets of characters, evaluation of the similarities and differences between them, and, in the case of phylogenetic taxonomy, inferences about the evolutionary history of the taxa derived from these comparisons. Yet such comparisons require corresponding or homologous characters for making the taxonomic judgments, regardless of the school of taxonomy followed. If we are to compare "apples and oranges," we must compare them over a set of characteristics applicable to both of them. In view of the crucial importance of the concept of homology in taxonomy it is indeed "... a serious flaw that the terms homology and analogy are so difficult to define, or at least their results are so difficult to distinguish" (Davis and Heywood, 1963, p. 40).

The concept of homology dates back to pre-Darwinian times. Owen defined a homologue as "the same organ in different animals under every variety of form and function," and he defined special homology (this is usually simply referred to as homology nowadays) as "the correspondency of a part or organ determined by its relative position and connections, with a part or organ in a different animal." He defined an analogue as "a part or an organ in one animal which has the same function as a part or organ in a different animal." Thus these terms had no necessary evolutionary connotations. To Darwin the existence of homology was another link in the chain of evidence for his theory of evolution. However, as soon as purported homologies were called upon to serve as evidence for phylogenetic relationships, any evidence concerning these homologies derived from the phylogenetic inference immediately became suspect. Taxonomists have wrestled with this problem to this day and critical examination has always shown a phylogenetic definition of homology to be wanting. Thus Simpson's definition that "homology is resemblance due to inheritance from a common ancestry" (1961, p. 78) is self-defeating, since it is a primary purpose of the phylogenetic school of systematics to work out phylogenies, and for this they need homologies that are not defined in terms of the conclusions they wish to reach. All of the criteria used by Simpson (1961, pp. 87-92) for recognizing homologies are subject to serious objections (the interested reader may find these in our *Principles of Numerical Taxonomy*). Hennig (1966, p. 93) defines homologous characters as those "... that are to be regarded as transforma-

tion stages of the same original character. . . 'Transformation' naturally refers to real historical processes of evolution. . ." Hennig realizes that this definition can only be of theoretical value, admitting "the impossibility of determining directly the essential criterion of homologous characters—their phylogenetic derivation. . ." He is forced to turn to Woodger's concept of correspondences (see ensuing discussion) in order to approach the problem of homology operationally.

Among the reactions to the impracticality of a phylogenetic concept of homology has been the school of idealistic morphology (for review and entry to the pertinent literature see Zangerl, 1948, Sokal, 1962b, and Jardine, 1967). The metaphysical form of this approach is rejected by most modern biologists, but empirical and statistical forms coincide with Woodger's proposals for a homology based on a theory of correspondences. This, as we shall see below, is an essentially phenetic concept. It is also noteworthy that chemists are increasingly using the term isology (introduced by Florkin, 1962), for chemical correspondence (for example the occurrence of cyanides in two organisms) rather than homology, when little is known about their evolutionary origins.

Before we turn to a detailed discussion of various operational means of determining homologies, we need to remove a misconception still current among many taxonomists: namely, that if only the fossil record were complete and known, phylogenetic classification and hence phylogenetic definitions of homologous characters would be possible. It has been pointed out by several recent workers that even with a fairly complete fossil record, the only way of associating specimens at the various time levels is by overall resemblance of organisms and their parts. Colless (1967b) has examined this question more fully than others. He postulates a completely competent and adequately long-lived observer who could continuously record the evolutionary history of a group of taxa as it unfolded before his eyes. He agrees that the hypothetical observer would clearly know the cladistic relationships among the various branches but would still have to measure overall resemblance to get an estimate of phenetic relationships among any points along the branches. Now suppose the observer examines the evolving group of taxa at regular time intervals. This makes the model slightly more realistic. The question now becomes whether at successive time intervals T_1 and T_2 he could perceive if a form A_1 was the same or had changed. He would recognize A_1 at time T_2 by its possession of the same complex of attributes it possessed at time T_1 . If the interval from T_1 to T_2 is substantial there might no longer be a form A_1 at time T_2 , but a hitherto unrecorded $A_{1,1}$, closely resembling A_1 , so that the observer unhesitatingly accepts it as a changed A_1 , that is, as an evolutionary descendant of A_1 . Colless (1967b) concludes: "The trend of this argument should by now be obvious. . . the point I am making is that *any* observer making surveys at T_1, T_2, \dots, T_n , can infer the course of evolution only from a detailed and direct comparison of the complexes of attributes possessed by taxa at each and every one of those times: i.e., by a survey of 'overall resemblances.'" Clearly then, even under impossibly

optimal conditions the determination of homologies in phylogenetic taxonomy is made essentially on a phenetic basis.

In the rest of this section we shall therefore concern ourselves with a phenetic and empirical concept of homology that we hope will become increasingly operational and quantifiable.

Homology Redefined

As we have seen, a precise and universally applicable definition of homology has not so far been proposed. Various definitions applicable to special cases will be featured below. By abstraction from them, homology may be loosely described as compositional and structural correspondence. By *compositional correspondence* we mean a qualitative resemblance in terms of biological or chemical constituents; by *structural correspondence* we refer to similarity in terms of (spatial or temporal) arrangement of parts, or in structure of biochemical pathways or in sequential arrangement of substances or organized structures.

In the past, homology has generally been a categorical concept, that is, structures either were, or were not, homologous. Attempts at making homology operational have generally led to a quantitative concept of homology that permits degrees of homology (e.g., Sattler, 1966). Thus, structures may be more homologous with one structure and less so with another. If the categorical definition of homology is preferred, then a character or a structure may be considered homologous with the character or structure with which it shares the greatest degree of similarity or correspondence, an approach Jardine (1967, 1969c) prefers. Compositional correspondence can be regarded as a special case of homology decided upon an external criterion, and structural correspondence as a special case of homology decided by internal criteria, as pointed out by Sneath (1969c). Jardine and Jardine (1969) prefer the terms "attributive matching" and "relational matching" respectively. In most biological situations the criteria can be adequately described as compositional or structural, which are the headings we employ for the present discussion, but sometimes this would unduly stretch the usual meaning of these terms. Thus the ability of a bacterium to ferment lactose would scarcely be called compositional (though it may depend ultimately on some unknown detail of composition), and it would be clearer to say that the character "lactose fermentation" is based on an external criterion involving a chemical test procedure. Indeed, in bacteriology the bulk of the characters employ external criteria. Sneath also points out that compositional criteria may introduce two additional aspects of homology. The correspondence may be quantitative, as when, in considering the correspondence based on the compositional criterion of calcification, we may find that some structures are more heavily calcified than others. Either the degree of calcification is admitted as a relevant quantitative variable or some level is used to separate calcified structures from noncalcified ones. The correspondence may

also be diffused, as when a bone has a correspondence to all other bones with respect to the criterion of calcification.

All cases of structural homology contain some external element in order to distinguish the parts under study from other parts (e.g., bones from soft tissues). However, the use of external criteria does not avoid all problems; thus we may be faced with a bone in one animal that is represented by a tendon in another, and Jardine and Jardine (1969) note that most cases of homology involve compromises between compositional and structural criteria.

Operational Homology

The first approach to homology in phenetic and numerical taxonomy was to call two character states the "same" whenever they are indistinguishable. Similarly, if the abstraction or idea of two characters cannot be distinguished in the taxa, then again we would consider them to be the "same." In practice the worker will divide his organisms into major structures or other such divisions (for example, head, limbs, leaves), and he must first decide whether these are the "same." Then he can proceed to look for differences in their properties that may be used as the bases for setting up the characters and their states. Within each major structure he again looks for subsidiary structures and repeats these procedures. Characters in this sense are synthesized from the states.

For example, consider two species of insects that are both black, while others in the same genus are red. If we had no way of distinguishing the two kinds of black, we would consider them to have the same character state, "black." Similarly, we would consider "redness" and "blackness" to be states of the same character, "body color," unless we had reason to believe that this color was of a different nature in some of the insects than in others. If, for instance, we found colors due to pigments as well as colors due to optical interference phenomena (such as iridescence) occurring in the group of insects under consideration, we would then subdivide our former character "body color" into two: "pigmental body color" and "structural (interference) body color."

By way of another example, leaves on a given series of plants may be long or short. We first have to decide what a leaf is and whether the structures seen on the separate specimens are in fact leaves—that is, are the "same"—or perhaps are other structures such as modified stems. Having decided that they are leaves, we also have to agree on what we shall call a short leaf—perhaps one of less than 3 cm: if so, we shall call a long leaf one that is longer than 3 cm. In this sense all leaves shorter than 3 cm are homologously short leaves, those longer than that are homologously long. The character will now be called "length of leaf," with two states, "short" and "long." Clearly it could have had more states had one wished to construct it in such a manner. In each leaf other characters, such as the venation, pubescence, and similar properties, could be coded as characters.

We call this approach *operational homology*. Our position is largely that of common sense; when we say that two characters are operationally homologous, we imply that they are very much alike in general and in particular. If the characters are "not quite the same," then more than one character is involved, and they should be broken into several independent ones; some of these independent characters will then be indistinguishable and will be scored as "the same" character in the two organisms.

A difficulty of the operational definition of homology is that some homology statements are made about unit characters, others about sets of unit characters. Some unit characters such as color, weight, length, or pH can be defined unambiguously in terms of operations necessary to ascertain their magnitude. These are usually properties of the entire organism, definable on external criteria. However, other unit characters are properties of parts of an organism, each part being defined by a set of unit characters. But these characters are in turn defined by the parts to which they belong and need reference to structural correspondences. Thus, for example, an eye or a forelimb can each be subdivided into numerous unit characters, yet each of these unit characters cannot be defined without reference to the fact that it is on the eye or the forelimb. For example, color of eye or color of appendage must be related to eye or forelimb. A circularity of comparison (Inglis, 1966) makes it difficult, therefore, to define an eye or a forelimb in any rigorous way. In practical work in numerical taxonomy such decisions are usually carried out on a common sense basis. However, for a rigorous definition of homology this is not sufficient. In fact even the entire organism needs to be defined unambiguously.

To establish that two eyes in different organisms are homologous, we can use two approaches. We can define them purely structurally, that is, consider an eye a unit character and describe its relation to other parts of the body (cranium, muscles, nerves) in such a way that it will become apparent that the characteristics in the two organisms are homologous; that is, both are an eye. Although this approach employs criteria external to the eye (e.g., a large nerve connected to the brain), it requires criteria internal to the organism and presupposes knowledge of the identity of these other parts of the body. Hence it depends on previous homologies established for them, finally resulting in the conclusion that the greatest number of identical relations is obtained if both are identified as eyes. The other approach is to use a criterion of similarity based on numerous finer unit characters, that is, statements about the components of the eye. By showing that the structure in question in one organism is most similar to the structure known as an eye in another organism, we can establish their homologies and identify the structure in question as an eye. The difficulty here is that in order to do so by the similarity approach, we must define unit characters in both structures, and such unit characters in theory cannot be defined unless we know that the structure is an eye. With both approaches, therefore, we are caught in a circularity of reasoning. Fortunately we can often make a start with compositional criteria such as presence of lens

protein, which in principle could be established on chemical grounds. A rigorous approach would be to consider the eye a unit character first, determine homology provisionally on a compositional and structural basis, and then proceed to the more detailed analysis involving sets of characters per organ or complex structure.

In calling this concept operational, we did not so much have the philosophic idea of operationism in mind (see Section 2.1). Rather, we felt that we could not rely on phylogenetic inferences; we had to have some procedure for getting started or "becoming operational." As we shall show later, when the approach of operational homology is taken to its logical foundations, it becomes a numerical homology and, while not fully operational in the sense of operationism, it is certainly considerably more so than the concept of phylogenetic homology. Hull (1968a) points out correctly that in determining homologies, pheneticists do not restrict themselves to direct observations of shape and relative position. In determining what is the same state and character in different organisms, some previous knowledge of the biology of the structures and organisms concerned is clearly employed. None of the procedures for finding homologies are entirely inference-free, and in order to bootstrap ourselves into a position from which we can make some statements we must initially assume certain relationships as being obvious. We may, however, then turn around and test these relationships. As the concepts of numerical homology are developed these inferences become less necessary, or at least more clearly defined. We cannot agree with Hull, however, when he equates these non-observational inferences in phenetic taxonomy with those in phylogenetic taxonomy, which he considers to be certain law statements concerning trends in evolutionary development, the relations between embryological and evolutionary development, the correlation between morphological and cladistic affinity, and the like. But here Hull has put his finger on the very weaknesses of this line of reasoning. There are few, if any, general "laws" that can be derived from any of the considerations mentioned.

The concept of operational homology as used by us is close to the original definition of homology as employed by Owen and discussed earlier. Woodger (1945) has discussed the problem of homology from the point of view of morphological correspondences, deriving the concept in what would now be called a phenetic sense, with structural and compositional considerations both contributing to the correspondences. He points out that in making such morphological correspondences we pair the different parts of the structure in two organisms, with the aim of obtaining the greatest number of one-to-one pairings. For example, if we pair head of cat with head of dog we find numerous subsidiary pairings within this major pairing—such as eye with eye, and brain with brain; further pairings occur within the latter structures—lens with lens, retina with retina, cerebellum with cerebellum, and so on, down to histological levels and farther. A similar comparison between head and leg would show few such correspondences. Woodger illustrates his arguments by the example of the pentadactyl limb. He shows that the pairing of the

“same” bone in two forms—humerus with humerus, radius with radius, and so on—depends on the spatial relations of the bones. For example, the humerus is proximal to the other bones, the radius and ulna are both immediately distal to the humerus, the ulna being postaxial to the radius, and so on. In a newly studied creature we call a bone the radius if we find that it bears these relations to the other bones, and if they in turn bear their own proper relations to each other. Such matching sets of bones Woodger calls *isomorphic*. He also discusses the difficulties that arise if some bones are atrophied or if their articulations are abnormal, so that it may be difficult to recognize which bone corresponds to which. Woodger correctly says that morphologists pair off organs one with the other intuitively so as to make the greatest number of one-to-one correspondences.

As we attempt to apply the ideas of operational homology in actual practice we encounter considerable difficulty. First, how are we to determine the homology of the large number of characters that are employed in numerical taxonomic studies? It is unrealistic to expect that in the near future methods will be developed for the rapid determination of the homologies of large numbers of characteristics. In almost all numerical taxonomic studies carried out to date, the judgment of the taxonomist regarding homologies is implicitly accepted. The homologies are deliberately phenetic, and evolutionary considerations in evaluating these homologies are intentionally avoided, although we cannot always be certain that this is done. However, since we believe that much purported phylogenetic homology is in fact a measure of phenetic and relational resemblances, some phylogenetic coloring of the judgments of the phenetic taxonomists is probably harmless. Operational homology, when carried out at this level, clearly does not conform to the dicta of operationism. Only as the concept of homology becomes more quantitative (see later discussion) will it be more operational in the strict sense.

Problems of homology have not been serious in numerical taxonomic work in practice. Fisher and Rohlf (1969) and Moss (1968b) found that numerical taxonomies are quite robust to small errors in homology. Fisher and Rohlf found with correlation coefficients that changes were minimal when less than 10 percent of characters were wrongly homologized (there was also an effect on distance coefficients, but this was smaller probably because the permutation process they used preserved the general size factor, see Section 4.11). It may be noted that in constructing phenetic classifications in bacteriology it is possible to formulate a list of characters and their states that does not presuppose a prior knowledge of homologies. Nevertheless these problems must be watched; Jardine and Sibson (1971, p. 171), for example, believe that a numerical taxonomic study by Hamann (1961) may be misleading because of difficulties in homology with families of monocotyledons.

To date, the various proposals for quantifying the concept of homology have sought to analyze the mental processes of systematists and comparative anatomists. These approaches can be classified under two main headings—the structural

approach in which correspondences in positional relations between parts are the primary criterion (but which also can include relations such as are exemplified in biochemical pathways), and the phenetic approaches, in which homology is based on overall similarities of organs and complex structures in terms of their constituent unit characters. In the various approaches below, the definitional approach applies to unit characters only. The other approaches all are based on sets of characters. *Structural correspondence* comprises (1) a multidimensional approach and (2) a one-dimensional approach. The *phenetic approaches* to homology include (1) the definitional approach, (2) similarity of complex structures based on defined unit characters, (3) similarity based on undefined characters, and (4) geometric similarity.

Structural Correspondence

Multidimensional Structural Correspondences. Much of the classical concept of homology as derived from the work of the comparative anatomists and especially as applied in vertebrate anatomy focuses on the positions of various parts relative to each other. The pioneer work in the rigorous formulation of this notion of homology is that of Woodger (1945), whose concept of homology involved 1:1 correspondences between morphological parts maintaining their positions relative to each other. These ideas have recently been further developed by Withers (1964). The idea of using evidence for homology from the position of structures is well established. Inglis (1966) states that in making judgments on homology from structural relations one must assume the known homologies of other structures in the organisms. "... Thus a statement of absolute resemblance involves a system of comparison, using as reference points as yet uncomparing organs or features, the validity of whose use in this comparison is only justified by later reference back to the first organs or features compared... Comparison leading to the recognition of absolute resemblance is therefore circular and involves whole organisms." Extending the line of argument, Key (1967) visualizes homology as being that state of arrangement or matching of characters that will result in the greatest estimated similarity between OTU's being compared. His definition of homology is: "Feature a_1 of organism A is said to be *homologous* with feature b_1 of organism B if comparison of a_1 and b_1 with each other, rather than with any third feature, is a necessary condition for minimising the overall difference between A and B." The difficulty with making such an approach practical is that in order to measure the overall difference between any two OTU's and minimize it by rearranging the characters, one must already have an acceptable system for coding the characters and to do so one needs to have at least a tentative concept of homology.

A significant breakthrough in quantifying the concept of homology has come from Jardine (1967). He follows Woodger in using the criterion of agreement in relative position and connections with each other. For the determination of

homologies, a set of relations is established of the type "anterior to," "ventral to," or "right of." A set of parts of an organism is said to be connected by these relationships, which must be nonreflexive, that is, if part 3 is anterior to part 1, part 1 cannot also be anterior to part 3. By Jardine's definition a *similarity* is a function rearranging the parts of one organism with respect to another, in order to obtain 1:1 correspondences with respect to the relations between the parts. That function which produces a unique maximal correspondence (in Jardine's term, a similarity having as members the greatest number of ordered pairs of parts) is called *h*, a *homology*.

These definitions lead to an algorithm that permits the computation of the unique largest similarity. This is the largest number of correspondences between some or all of the parts from each organism in which the spatial relations between parts are preserved. Jardine's program permits "undecidable cases" and the program has the option of allowing for two or more parts to be considered together as a single part. This permits a part in one organism to correspond to an aggregate of several parts in another. Jardine and Jardine (1967) have applied this technique to a number of simple cases such as bones of human, cat, rat, and dog skulls, where it yielded satisfactory results. In investigating a more challenging case, the homology of the skull bones of the holostean *Amia calva* and of the teleost *Clupea finta*, the method did not yield any unique largest correspondence but several equally large correspondences. These differed in the way in which the bones of the suborbital series were paired but gave consistent correspondences between other bones. The results largely supported W. K. Gregory's interpretation of these homologies.

Despite the success of his approach, Jardine (1969a,c) has shown that different early assumptions—expressed in the list of relations admitted as relevant—are critical. Thus, in computing the homology of the skulls of the cat and rat, the generally accepted homology is found if the relations "anterior to," "ventral to," and "distal to" are employed. Moreover, if the relation "adjacent to" is added, several equally large correspondences are found. This is because the frontal and squamosal bones in the rat are adjacent, unlike the cat. In a detailed study of skulls of fossil fishes Jardine showed how different assumptions of this kind would give maximal correspondences that agreed with the homologies proposed by various authors, each of whom had apparently made different assumptions of this type. This was particularly noticeable with decisions on whether the lateral lines (sensory canals) could move from one bone to another during evolution. Difficult problems are also posed in distinguishing between the fusion of two bones and the loss of one of them.

Such observations reveal our inadequate knowledge of how to reach appropriate assumptions. Although Jardine is undoubtedly correct in saying that a background knowledge from fossil studies plays a part in this (for example, bones rarely change their antero-posterior relations, but commonly change in adjacency) we would not

overemphasize this dependence on paleontology. If this dependence were critical it would prevent homologization in groups without fossils, and a child who knew nothing of phylogeny could not homologize the hand of a man with that of an ape.

Jardine's approach is clearly a numerical homology, but as contrasted with those efforts based on phenetic approaches, his measure of similarity is maximum correspondence within a network of relationships. Jardine's approach is a considerable advance over earlier more or less intuitive assessments of relational homology. As formulated, it clearly determines the homology of unit characters (the parts). The extent to which it can be applied for characters in general is still in question. First, though it can be applied to subunits of parts in terms of its logical and mathematical formulation, can we make relational statements about ever-smaller parts or bones in the same way we can about discrete bones or sclerites? Here is where the compositional aspect of homology (cartilage, connective tissue, muscular tissue, etc.) could usefully be combined with the structural relation. Also, more complex relational statements will be made as we extend our purview beyond strict classical morphology. Can such a system be employed for homologizing biochemical pathways or biochemical structures? What of behavioral homologies? Possibly a "hybrid" system employing phenetic similarities between complex structures as well as the greatest numbers of relational correspondences may be developed.

One-dimensional Structural Correspondences. In biology this approach to homology has mainly concerned amino acid sequences in proteins. Whether approaches of this sort might be useful in other sequentially organized structures or properties, such as metameric organs or organisms, remains to be investigated. One of the ways of finding homologous positions in protein chains is to align the chains side by side in the manner that gives maximum agreement between pairs of amino acids. This is the method for homologizing amino acid sequences by cross-association (sliding matches) as described by Sackin, Sneath, and Merriam (1965). A second method suggested by Fitch (1966a) finds the minimum number of mutational changes in the nucleic acid sequence that would convert one protein into another. Thus, if at a given position in the chain one amino acid is glycine and the other phenylalanine, the minimum change would be two nucleotide mutations, from glycine codons GGU or GGC to UUU or UUC (the only codons for phenylalanine), and not from the glycine codons GGA or GGG, which would require three mutations to give phenylalanine. By arranging protein sequences according to their best matches, homologous positions may be established and evolution at these sites investigated. The complexities and difficulties in homologizing protein sequence positions when there are deletions or gaps in one chain (see Cantor, 1968; Sackin, 1969, 1971; Fitch, 1969, 1970a,b) are such that even the simpler sequential correspondences raise unsuspected problems.

Phenetic Approaches

The Definitional Phenetic Approach. This is especially suited to unit characters that are not part of an organized structural tissue. Identically defined characters can be described in different organisms; these are homologous by definition. This definition may be in terms of general appearance, structure, or composition. Thus, weight or maximum length, if rigorously defined, is homologous by definition in two OTU's. Similarly, the presence of hemocyanin in organisms would be homologous by definition. Note that these statements are not related to any organ or tissue and thus need no inferences about previous homologies of such structures.

Similarity of Complex Structures Based on Defined Unit Characters. One way to establish homology between organs, organ systems, and other complex structures is by their overall similarity. This process is comparable to the classification of the organisms themselves. It attempts to define structures in terms of unit characters (Section 3.3) and applies to them the concept of operational homology discussed above. Some intuitive judgment enters in such definitions of characters. The aim is to compare these organs, character by character, and ultimately to group structures so that there is the greatest number of common properties in the "organ taxa" set up. We may call those structures homologous that exhibit maximal similarity with each other among a set of potential structures derived from the two organisms being compared. This is evident from a consideration of why we pair an eye with another eye and not with an ear. In both alternatives there are some shared character states (both organs are carried on the cranium, both are special sense organs), and some differences (no eye is identical with any other eye), but there is no single property which we can satisfactorily postulate a priori as being essential to the definition of an eye or an ear (since we may be sure that some morphologists will soon find an exception to our rule). Therefore, the concept of natural organ taxa is polythetic (see Section 2.2), just as natural taxonomic groups are; hence natural organ taxa can, like natural taxa, only be defined as arrangements by which the groups so formed possess the greatest possible number of common properties. Such reasoning leads us to a *numerical homology*—that is, a numerical taxonomy of organic structure. As pointed out by Colless (1969a) this would also lead further to the homologizing of the whole organism, and numerical homology would then be intimately related to some aspects of phenetic resemblance. In contrast, the definitional approach is analogous to monothetic grouping methods.

By basing homology on maximal overall similarity and basing the latter on correspondences between unit characters, we find ourselves with the realization that increasing knowledge and refinement of techniques will lead to ever finer unit characters in terms of structural complexity. This in turn leads to clusters and structures at increasingly finer levels. Thus not only do we find within the concept

of natural taxonomic groups similar concepts of natural organ groups, but also within these there may be yet other like concepts (such as natural cell groups or natural gene groups) hierarchically arranged like a nest of Chinese boxes. There must, however, be some limit to this process, even if the limit lies at the fine structure of the genes.

However, even if we proceed to this level, the problem of homology remains with us, now commonly called *genetic homology*. Eventually one may visualize genetic homology as correspondences in the nucleotide base sequence of the double helix of DNA. At this stage in our understanding of molecular biology it may be more useful to think of amino acid sequences of proteins. It is possible to identify regions of the protein molecule in which corresponding sequences can be found even in quite distantly related organisms. For example, cytochrome *c* seems to have regions that can be traced back to a common ancestor of yeast and man.

There are still many problems in any approach to genetic homology. The sporadic occurrence of enzymes unusual in animals, such as the carbohydrate-splitting enzymes of the snail, is presumably based on evolution *de novo*, for it would be difficult to believe that these enzymes in snails are derived from their remote protistan ancestry. It seems very likely, however, that when such re-evolution does occur it is due to reduplication of a gene controlling the production of a similar enzyme, followed by change in the enzymatic properties of one of the duplicate enzymes. We assume that the actual sequence of amino acids is duplicated and leads to separate evolutionary structural divergence. The classical instance of such a phenomenon is the reduplication of hemoglobin genes that have undergone independent evolution in mammals (Ingram, 1961), possibly with internal duplication as well (Zuckerandl and Pauling, 1965b; Fitch, 1966b). A further possibility is the transfer of genes from very dissimilar organisms, by various mechanisms. For example, viruses might transfer genes to distantly related plants or animals (Sokal and Sneath, 1963, p. 73; Anderson, 1970).

An important consideration with reference to tracing a homology to the levels of fine genetic structure is whether we should, in fact, do so. This problem is discussed in greater detail in Section 3.6 and here we might simply point out that for many types of evolutionary and taxonomic problems it would not be desirable to proceed to the molecular level, but to evaluate homologies at a higher structural level.

Considering homology as maximal overall similarity between complex structures also has implications for the definition of *analogy*. We have not so far paid much attention to definition of this concept. If by analogy we mean functional but not morphological similarity, one must be able to separate logically the definition of structure and function, which may not be easy. It seems to us that function is also an aspect of the phenotype and that statements about function can be unit characters in the same way that statements about morphology can. If that point of view is accepted, then it follows that functional identity or similarity will lead

to phenetic similarity just as morphological identity does. Furthermore, if overall resemblance is based only on functional characters, it may well be that structures will be considered to be highly similar that, in fact, might have considerable morphological disparity. When estimating overall similarity we should not separate structure from function; we would not wish to base similarity simply on function (unless we intend to make a special classification such as one for functional anatomy or for ecology). We do not believe that false overall similarity (due to analogies) would enter very prominently or frequently into a phenetic measure of homology since analogous organs (wings of birds and butterflies) would be dissimilar in so many structural and functional character states compared to their few common functional and morphological character states. So considered, homology and analogy are not categorically distinct concepts, but are near one end point and at an intermediate point, respectively, of a continuum of relationships in which total identity is at one extreme and total dissimilarity at the other. The general region of high similarity is homology. Cases of analogy would be found at intermediate levels of similarity, where it could be shown that the subset of characters dealing with functional relationships expresses similarity though a subset of morphological characters does not. Rather than make a categorical contrast between homology and analogy, therefore, we prefer to speak of greater or lesser resemblance among organs, organ systems, and other levels of complexity.

Similarity Based on Undefined Characters. One way to deal with problems of homology is to ignore details of structure, given certain preliminary assumptions about the spatial orientation of specimens with respect to the observer. This is the approach undertaken experimentally in a study by Rohlf and Sokal (1967). Basically, the technique consists of recording agreement in visible structures over selected minute areas of the images of two organisms. This was accomplished by preparing random masks from 25 punch cards, each perforated with 25 randomly chosen holes. These masks were placed over black and white drawings of 29 "species" of Caminalcules (a group of imaginary animals created by Joseph H. Camin; see Section 6.4) and 32 species of culicine mosquitos. Black lines appearing through a hole were scored 1, and empty holes were scored 0 (see Figure 3-1). Similarity matrices were constructed on the basis of matching scores for corresponding masks and holes (625 pseudocharacters). The taxonomic structure obtained from scanning images in this way compared favorably with the structure obtained by deliberately looking for and scoring characters. The surprising results would suggest that precise homologies may not be necessary for expressing phenetic similarity between organisms. In fact, however, correspondence between scans does rest on general morphological similarity both in structure and in relationship of the images compared. The "characters" in this study are thus not biologically (morphologically or relationally) defined but represent given positions in the scan. It is quite possible that in complicated images parts of the image might be demonstrated homologous

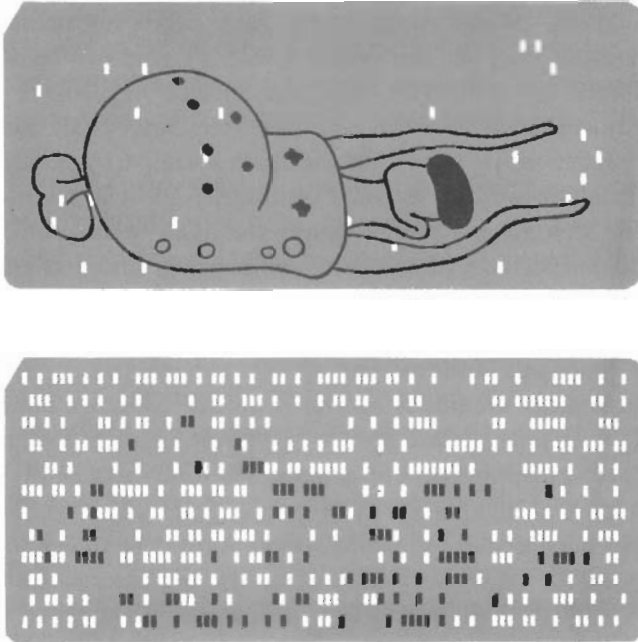


FIGURE 3-1

Automatic scanning of Caminalcules. At the top, one of the 25 masks, containing 25 randomly chosen holes, is shown superimposed over a drawing of a Caminalcule. In making the composite shown below, each of the 625 holes was scored 1 when a black line appeared through it and a 0 when no black showed. (Actually, fewer than 625 holes are visible here, since many holes on different cards coincided). [From "Numerical Taxonomy" by Robert R. Sokal. Copyright © 1966 by Scientific American, Inc. All rights reserved.]

by the scanning method. Complete scans of the entire image (all positions on the grid or mask as contrasted with the random scans reported above) also gave satisfactory results, but classifications were not necessarily better than those achieved by the random scanning method.

This approach has clear limitations. Great disparity in size of organisms would lead to unjustified dissimilarity, so that adjustment to a common scale would be essential. In their work with the Caminalcules, Rohlf and Sokal made no attempt to allow for differences in size. Thus a large proportion of the phenetic differences recorded by the random scanning method were simply differences in size. In spite of this introduction of obvious error (unless size should be weighted more heavily than is usually done by biologists), the taxonomic structure obtained was quite satisfactory. In aligning the organisms before scanning, a primitive type of homology is used. Decision must be taken on what is anterior and posterior in the animal, the location of the "shoulder," bilateral symmetry, and so forth. The orientation of the images is clearly important, and consistent criteria must be chosen.

Differences in the orientation of appendages, such as wings or legs, would make related organisms appear far more dissimilar than they otherwise would be. Such adjustments could be made by operators on graphic display consoles or automatically by rather sophisticated programming (as described below).

Geometric Similarity. Rather than estimate the similarity between samples of images as just discussed, one could try to fit entire images to each other. This would require enlargement, diminution, rotation, translation, and disproportionate stretching of some images to a standard or pattern image. Such an approach could in theory be undertaken entirely "blindly." By this we mean that an attempt could be made to fit the outline of image *B* to that of image *A* in the best possible manner. Even if these images are not at all phenetically related, as for example the outline of a *Paramecium* and that of a human skull, there will nevertheless be some position of best fit. The goodness of this fit should be of a low order. When homologous images are compared, these differences should be much less and it should be possible to transform one image essentially into the other.

If one employs the method of matching diagrams described by Sneath (1967a), it is necessary to mark homologous (i.e., corresponding) points on the diagrams. This can be done by eye, in most cases no doubt without difficulty. Thus we might mark the tip of the nasal bone on two skulls. This involves decisions on homologies, for example, what is the nasal bone and what is its tip. Sneath notes that one can envisage an automatic method, though it has difficulties that have not been explored, which would represent the lines of the diagrams as series of closely-spaced points and scale the figures to the same size, overlapping them at their centroids. One diagram, *B*, of a pair under comparison (*A* and *B*) is then rotated, so that the sum of distances (or squared distances) between every point on *B* and the nearest point on *A* is minimized. There is no analytical solution to this, because radially symmetrical objects would show several positions of low misfit (local minima) and the lowest of these would have to be found (the global minimum). This would require a succession of small angular rotations to find the global minimum, which should, as a rule, lead to the correct general orientation of the two figures (both skulls for example, facing the same way). There then remains the problem of fitting smaller areas of the diagrams to obtain homologies between the finer details, which might be carried out successively for smaller and smaller areas. In theory the end result of this process would give progressive distortions of the parts of the diagrams until every point on diagram *B* lay on top of the homologous point in *A*. The mathematics of performing the distortions would pose considerable difficulties in computation, though trend surfacing of the displacements between the points of *B* and the nearest points of *A* might offer a solution.

It is not yet clear which of the above approaches will be the most fruitful in having the greatest general application for taxonomy. Quite possibly an integrated approach employing several of these concepts will be developed. In the meantime,

the numerical taxonomist wishing to proceed expeditiously in classifying groups of organisms may need to rely almost entirely on his intuitive appreciation of homologies when defining and coding characters. Other studies undoubtedly will be carried out whose main aim will be establishing homologies between characters or organized sets of characters. All such studies, regardless of the approach employed, if they are eventually to yield data matrices for numerical taxonomy, must employ some form of maximal criterion of similarity or correspondence to establish the homologies; in effect one must use a categorical concept of homology.

3.5 KINDS OF CHARACTERS

The philosophical may argue that it is not possible to make absolute measures of resemblance, because such measures would involve an arbitrary selection among the endless array of attributes which could in some sense be called characters of the organisms. Nevertheless, meaningful estimates of resemblance can be made once there is agreement on what characters are to be admitted as relevant in taxonomy.

It is, of course, quite impossible to give an adequate catalog of all the taxonomic characters that can be used in various groups. Such a catalog would comprise nothing less than a description of organized nature. Only specialists in the various groups will be in a position to define and describe unit taxonomic characters in the organisms they are studying. In their search for characters they ought to follow two guide lines, one of which at least is not included in customary taxonomic practice. First, use all kinds of characters from all parts of the body and from all the stages of the life cycle. Second, use all characters varying within the group studied, not merely conventional diagnostic characters. The latter are likely to be constant within the members of a given taxon. The exclusive use of such characters would prejudice the very issue—the establishment of taxa free from subjective bias—which numerical taxonomy wishes to solve. If the studies are based to a very large degree on characters previously described in the literature, there is some danger that diagnostic characters will be favored inordinately, since there is a historical weight in favor of diagnostic characters in the published literature. There is at least the suggestion of such a bias (Sokal and Rohlf, 1970) in the original set of characters describing the *Hoplitis* complex (Michener and Sokal, 1957), chosen by Michener from his published work and his recollection of the taxonomic relationships of this group. Reanalysis of the same data by various techniques has shown that the Michener and Sokal data matrix yields tighter clusters with wider gaps and fewer intermediate species than comparable taxometric studies of other organisms.

Taxonomic characters can be grouped roughly into

- a, morphological characters (external, internal, microscopic, including cytological and developmental characters),
- b, physiological and chemical characters,
- c, behavioral characters,

d, ecological and distributional characters (habitats, food, hosts, parasites, population dynamics, geographical distribution).

This list is far from being exhaustive. Readers interested in greater detail may wish to consult Mayr (1969b, p. 127), who has prepared a more complete classification and discusses each category in detail. Davis and Heywood (1963) may be consulted on characters in botany and Skerman (1967) and Lockhart and Liston (1970) for bacteriology.

From the above list of characters we single out chemical characters in group *b* and ecological and distributional characters (group *d*) as requiring special discussion.

In recent years there has been a great increase in work on chemical constituents as taxonomic characters. Comprehensive references to the literature are given in Blackwelder (1967a, p. 624) for animals, and by Hegenauer (1962–69) and B. L. Turner (1969) for plants, and a recent symposium (Sibley, 1969) covers all these groups. There are possible applications, too, to paleontological material (Abelson, 1957; Florin, 1966). Reference was made in *Principles of Numerical Taxonomy* to techniques that yield many characters in a single technical procedure, called polyphenic methods. These include chromatography, electrophoresis, spectroscopy, and other techniques; they are now widely used and are reviewed in the references cited above.

Although they may pose certain unusual problems in coding, *chemical characters* are just as valid for taxonomy as any others. Despite a few claims to the contrary, they seem on the whole no better and no worse than any other kind of character for constructing natural taxa, discriminating between species, or other systematic work. There are, however, two points about them that are apt to be misunderstood. First, there is a prevalent tendency to assess the chemical resemblances intuitively, and to compare chromatograms and the like by eye. Even when chemical resemblances are estimated directly, there is often no awareness that the next, and necessary, step is to search for taxonomic structure (for example by cluster analysis). We discuss numerical taxonomic methods for chemical data in Section 5.12.

Second, there are two main classes of chemical information which, although not always sharply distinct, have very different implications in taxonomy. Simple chemical substances, enzymatic reactions, and some other sources of chemical data yield relatively little information about the organism. They may be dependent on a few metabolic properties and are commonly recorded as present or absent, or given in quantitative terms. These are what Zuckerkandl and Pauling (1965a) call "episemantic molecules." Other classes of biochemical compounds are quite different in that their structures are highly complex, and, if known, yield a great deal of information. The nucleic acid coding of the genome is one example, and protein sequences is another: these are "semantic molecules" or "informational macromolecules." For this reason, methods that rely on the high information content of a protein, for example, are far more informative taxonomically than the presence

or absence of a single simple chemical compound. These methods comprise protein sequence studies, serology, and nucleic acid pairing, and they are discussed below.

Crick (1958) foretold the advent of *protein taxonomy*, which is now developing swiftly. A compendium of present knowledge is the *Atlas* of Dayhoff (1969a), and a good simple account is given by Dayhoff (1969b). Functional proteins commonly consist of two or more different sorts of polypeptides held together by various forms of chemical bonding. Thus adult human hemoglobin consists of two Hb_{α} chains and two Hb_{β} chains, forming a tetramer. Most interest for our purpose lies in the sequence of amino acids in a given polypeptide. The human Hb_{α} chain contains 141 amino acids, and the sequence starts, Valine-Leucine-Serine . . . , and ends, . . . Phenylalanine-Leucine-Alanine. There could be an enormous number of different polypeptides of length 141. In general there will be 20^n different polypeptides of length n .

The sequence of amino acids differs in the same functional protein of different animals. Thus human and horse cytochrome *c* differ in 12 out of the 104 amino acids. Protein sequence studies also have a unique power not possessed by any other taxonomic method; they allow estimation of resemblance between organisms that are exceedingly diverse. It would, for example, be hard to think of any reliable characters to use to compare man with yeast (morphology, simple chemicals, cytology, etc.). Yet the protein sequences of the cytochrome *c* of man and yeast are identical in 62 out of the 99 comparable positions, while the expected number for random sequences of amino acids is only about 7; indeed, it is even possible to detect some resemblance between cytochrome *c* of bacteria and man (Sackin, 1969). The number of identities and nonidentities between amino acids is therefore of great potential interest in many areas of systematics. The problem of character selection does not arise in its ordinary form (Section 3.6), for the sequence contains all the characters the protein possesses, at least if the primary structure alone is considered.

For a given kind of protein, such as hemoglobin, it is already evident that relatively small parts of the molecule have a fixed structure (i.e., in all organisms), presumably because of functional requirements. The folded structure of a protein appears to be determined principally if not entirely by the sequence of amino acids in the polypeptide chains. This sequence in turn is determined by the nucleotide sequence in the messenger RNA, and that is determined by the nucleotide sequence of the DNA.

It must, of course, be pointed out that the structure of one protein represents the fine structure of only one cistron. Whether or not the differences in any one protein are representative of the differences in the whole genome is a question that is taken up in more detail on the next pages and in Section 5.12, but the concordance is sufficient to make protein sequences a particularly interesting new field.

None of the chemical techniques discussed so far yields measures of taxonomic resemblance directly, though by counting similarities and differences in chemical

substances they can be made to do so. The next two techniques discussed in this section are different and they give resemblance measures directly.

A technique rather different from those considered so far is *comparative serology*. It yields quantitative measures of taxonomic relationship, but does so through the production of antibodies in experimental animals (usually rabbits). The basic principle is of wide application: a protein of one organism will react strongly with antibodies against it, but the same protein from a different organism will react less strongly. For example, antibodies to horse albumin, made by injecting the albumin into a rabbit, will react strongly with horse albumin, but weakly if at all with albumin of cow or pig. These reactions can readily be quantified, and are well known to be as a rule congruent with phenetic similarity. The characters that determine these reactions are usually not known, for they consist of details (as yet poorly understood) of the structure of biological macromolecules, usually proteins. General considerations of the nature of these characters are taken up here, and the kind of relationship that is revealed by serological techniques is discussed in Section 5.12. General reviews of this field are given in Leone (1964) and Hawkes (1968).

The many small differences and resemblances in proteins can be thought of as a large sample of the features of the organism. The fine structure of the genes in the form of the nucleotide sequence is translated into the sequence of amino acids in the proteins, as has been discussed. This, in turn, is expressed in the serological reactions and is one of the major factors determining these (e.g., Arnheim, Prager, and Wilson, 1969). On the other hand it is clear that the sequence of one protein cannot be taken as a random sample of all the genetic features of the organism. There is, therefore, a danger that if we study serologically a single protein (this is desirable for technical reasons), we may in effect be studying the fine structure of only one gene, which may not be representative of all the genes. It may be noted in passing that the value of serology and protein sequence studies does not depend on the "conservatism" of the proteins, for this is not what is required either for phenetic or cladistic studies; if all birds had retained the same serum proteins as their reptilian ancestor, then avian serology would be uniformly uninteresting. In fact, the information we now have shows that evolution in protein sequences is quite congruent with evolution in other respects, and also with phenetic taxonomic relations. This is supported by other evidence; the well-known correspondence between orthodox taxonomy and serology is one piece of evidence; another is the congruence between serological resemblances based on different kinds of proteins (albumins, globulins) and between proteins of different stages of the life cycle (Boyden, DeFalco, and Gemeroy, 1951; Wilhelmi, 1940; Spiegel, 1960; Marable and Glenn, 1964; Mohaghehpour and Leone, 1969).

At present we still do not have very much evidence on whether the many protein sequence differences have approximately equal weight in determining the serological results, but it is likely that this is true to a certain extent, although some

physicochemical features are essential to antigenicity. Therefore there is good reason to suppose that comparative serology should generally yield the same conclusions as numerical taxonomy, provided the sequence differences of the proteins used for serology are fairly representative of the whole genome. One might speak of it as a method for estimating resemblance in which the immunized animal acts as the computing machine when it produces the specific antibodies.

It may sometimes be possible to break down serological data to give antigenic formulas for the different antigenic factors. Where this can be done, these factors can be included like other characters in numerical taxonomic analyses (Lockhart and Holt, 1964), but more often the serological similarities must be treated as resemblance coefficients (see Section 5.12).

There seems no likelihood that any of the methods mentioned above will prove to be an adequate *sole* basis for taxonomy. To qualify as such, a method would have to reflect accurately the entire phenome of a given level of complexity (see Section 3.6). The last of the chemical methods discussed here—that using nucleic acid pairing—attempts to do just that; its future development will therefore be watched with great interest.

The *nucleic acid pairing technique* (often called hybridization) was pioneered by Doty, Marmur, Eigner, and Schildkraut (1960), and technical advances were made particularly by McCarthy and Bolton (1963). The “characters” are the nucleotide bases of the nucleic acids (most often the DNA of the genome) in their specific sequences. The resemblances thus obtained are discussed in Section 5.12.

Biochemists have sometimes argued that their characters are superior to morphological ones because they are closer to the genotype, and that chemical data can be arrayed in the order: DNA pairing, protein sequences, serology, and simple chemicals as gene products; each is increasingly removed from the nucleotide sequence of the genome. This argument is attractive on theoretical grounds, but we do not believe it should be emphasized much in practice. We still know little about the fidelity of expression of the genotype in chemical or morphological systems; technical problems (including the choice of resemblance measures) can vitiate the results and the adequacy of the techniques can only be assessed against other phenetic criteria; the biochemist may often only sample very small parts of the genome. Also, as we note in Section 3.6, classifications of genomes may not necessarily be the ideal of all systematics.

The factors of *ecology* and *distribution*, which have become much emphasized in recent years, require some comment in connection with their use for numerical taxonomy. While they are regularly reported, when known, they are not too frequently employed for classificatory purposes. Some difficulty may be encountered in coding them for numerical taxonomy. When an ecological character expresses some sort of gradient, such as life zones in mountainous areas, depth in soil, temperature maxima, and other gradients, multistate coding is straightforward. However, how are we to code phytogeographic or zoogeographic distributional

characters, host plant preferences, or parasitic fauna? With distributional data, a two-dimensional breakdown into two characters is sometimes possible. In the other cases the information may have to be partitioned into a number of two-state characters. Thus, where several species or genera of hosts occur for a group of parasites under study, each of the former may have to be a single two-state character marked "present" when parasitized, and "absent" when not. We may, however, wish to express systematic relations among host plants by appropriate coding. If a group of parasites lives on four hosts, species **A** and **B** of genus **X** and species **C** and **D** of genus **Y**, we could have one character for genus **X** or **Y** and one for species (**A** or **B** and **C** or **D**) for each of the two genera.

Geographic distributions are characters that may need to be used with caution. In most cases it is not possible to be sure that they represent any character in the genotype. Similar care is needed with many ecological, behavioral, and parasitological observations. For example, parasitological characters may sometimes depend on chance infestation and not on the genotype of the host or of the parasite. Thus, attempts to use the Mallophaga in classifying birds, for example, are made difficult by doubts as to whether some mallophagans are stragglers from other birds or have quite recently become established on their hosts through cross-infestation (for example, see Clay, 1949, and the discussion of Hennig, 1966 pp. 107, 175). Sometimes, however, there is remarkable congruence between phenetics of hosts and parasites, and Kistner and Pasteels (1970) describe an instance. Characters such as host specificity also pose the difficulty of a decision about how they should be scored. Thus some viruses are restricted to one species of mammal, while others attack both birds and mammals. One might decide to score class specificity, ordinal specificity, familial specificity, generic specificity, and so on, and to consider that higher categories should contribute more weight than lower ones. Since we have little detailed information on this problem, and hence no satisfactory method of allocating weights, we would suggest that only a few such characters be used and that each should be given equal weight. Although this may reduce the information, it will also avoid introducing bias. Ibrahim and Threlfall (1966a) give an example of such a study on fungi. Rather similar work has been done by 't Mannelje (1967b, 1969), and by Colwell, Moffett, and Sutton (1968), where susceptibility to the root nodule bacterium *Rhizobium* and to bacteriophages, respectively, were used as characters for numerical taxonomy. As we note in the next section the question of whether characters are genetically determined may not be so important as has been thought.

Little need be said about avoidance of bias in choosing the characters. It is clear that when we use only a set of characters known to show resemblance between certain groups, the similarity coefficients that will result from this study will reflect that choice. In an extreme case, by choosing only those characters that were the same in two organisms, one would obtain perfect but spurious resemblance between them. A systematic survey of all known characters, or the inclusion of all characters

the investigator has been able to observe, should prevent bias of this sort. Almost every new technique in biology gives new characters that can be employed in systematics. These new characters must be incorporated into the existing body of taxonomic data, and it is our belief that only numerical taxonomy can adequately do this.

3.6 CHOICE OF CHARACTERS

Having defined unit characters in taxonomy in Sections 3.2 and 3.3, and having surveyed the kinds of taxonomic characters in Section 3.5, we now proceed to the difficult question of which characters should be chosen as a basis for estimating the similarity between OTU's. In attempting to answer this question, we encounter two relevant problems: (1) What biological factors do the characters represent? (2) Are all characters of equal value and information in providing evidence on phenetic similarity?

Implied in the first problem was the hope during the early work on numerical taxonomy that taxonomic characters would provide information on the genetic factors differentiating OTU's. Although numerical taxonomy measures similarity between the *phenomes* of OTU's (Soulé, 1967b; defined by us as *the total phenetic manifestations of the genome of an organism or a taxon*), it was hoped that this would also lead to estimates of similarity between the genomes of these OTU's. In this connection we developed the *nexus hypothesis* which assumes that every phenetic character is likely to be affected by more than one genetic factor and that, conversely, most genes affect more than one character. The result is a complicated nexus of cause and effect. Any character should give information about several genes and it should be possible in general to pick up the effect of a given gene through any one of several characters.

While the nexus hypothesis is undoubtedly true in a general way (for detailed evidence see *Principles of Numerical Taxonomy* or general references to physiological and developmental genetics), we have recently felt that it has lost some of its relevance for phenetic taxonomy. The distinction between phenotype and genotype has become rather vague in view of recent insights into the nature of fine genetic structure. We have already suggested in Section 3.4 that the complexities—at many levels—involved in translating the genetic code into the final elaborations of external and internal morphology, physiology, and behavior are of such an order of magnitude that it is probably futile to hope that these can be understood in the near future. Not only are these interactions complex, but they also involve switch-mechanisms so that part of the genome of an organism is not functional at any one life history stage. What one wishes to measure in phenetic taxonomy is the *expression of the genome* of the organism through its life history—its phenome, in fact. Realizing this, it becomes less important from our point of view to know to what degree the phenome reflects the genome. We remain saddled with the problem of

having to investigate different levels of organized complexity, but for a general taxonomy all known levels should be utilized. For special classifications, external characters or biochemical ones or others might be preferred.

In discussing which characters to employ in a given taxonomic study, we earlier postulated the *hypothesis of nonspecificity*, which stated that there are no distinct large *classes* of genes affecting exclusively one class of characters, such as morphological, physiological, or ethological characters, or affecting special regions of the organisms, such as head, skeleton, or leaves. If this assumption were warranted, then obtaining a disproportionately large number of characters from one body region or of one special kind would not restrict our information to a special class of genes. Furthermore, there would be no a priori grounds for favoring one character over another.

The nonspecificity hypothesis has been shown to be only partially correct. Identical classifications are not produced from different sets of characters for the same OTU's. Measurement of the agreement between two classifications can most simply be done by correlating the matrix of similarity coefficients based on one set of characters with the matrix based on the other set of characters. This is one version of the method of cophenetic correlations, which we now prefer to call matrix correlations (Section 5.10). The resulting coefficient $r_{S_1S_2}$ between two similarity matrices (or, occasionally, $r_{C_1C_2}$ between two phenograms) estimates the congruence between classifications implied by the similarity matrices (or phenograms). By *congruence* we mean the *degree of correspondence between arrangements of OTU's in a classification*. Identical classifications are perfectly congruent. Tests of congruence may be made at a variety of organizational levels. Thus, we may look for the congruence between classifications based on different organs or regions of the body. Examples would be classifications based on characters of the brain contrasted with those based on characters of the intestinal tract; or characters of the epithelium contrasted with those of the connective tissue; or head versus body or wing characters. Other measures of congruence could be for dimorphic or polymorphic manifestations of an OTU, such as between classifications based on females or males, or those based on diverse life history stages, such as larvae, pupae, and adults in insects, or different castes in social insects, or different adult forms in cyclomorphic organisms such as aphids.

On the whole, classifications based on separate sets of characters—be these from organs or life history stages—agree partially. In vertebrate paleontology and systematics of recent vertebrates, classifications based on skeletons—or even portions of skeletons—are frequently congruent with those based on other parts of the organism's anatomy. It is well known that morphological evidence from a newly investigated organ system frequently confirms previous classifications.

Although the problems of incongruence have been fully appreciated by taxonomists for a good many years (Hennig, 1950, and Remane, 1956 give good reviews of this), it has only been through numerical taxonomy that quantitative estimates

of congruence have become possible. Table 3-1 summarizes a selection of measures of congruence obtained to date, listing the organisms, the numbers and types of characters, and the measures of congruence expressed as matrix correlation coefficients. It will be noted that there is a considerable range of these coefficients, from quite low values (classifications essentially independent) to fairly high ones.

Rohlf (1965) has applied a randomization test, such as are described in Sokal and Rohlf (1969), to test the significance of matrix correlations for three sets of data reported in Table 3-1. Rohlf's results indicate that correlation coefficients between matrices based on subsets of characters from different body regions or life history stages (with the exception of that between distance matrices in the butterfly data by Ehrlich and Ehrlich, 1967) are significantly lower than a random allocation of characters to each of two subsets would produce. Similar findings were obtained in a preliminary study of this point (Sokal and Sneath, 1963, p. 89; and by others). It is therefore unlikely that the subsets of characters that have been compared in these measures of congruence described identical taxonomic relationships. Readers will notice that the estimate of the congruence is in part dependent on the kind of similarity coefficient (correlation, distance, or other coefficient) on which the matrix correlation is based. This aspect of the matter does not relate directly to the problem of congruence but to the different nature of these coefficients, which will be discussed in Section 4.7.

On theoretical grounds, as well as from the results obtained so far, it would appear that congruence will be greater the higher the rank. Thus, for example, there is little doubt that the orders of insects would be as faithfully reflected in their larvae as in their adults, or that the classes of vertebrates would be recognizable from their skulls, their pelvic girdles, or their circulatory systems, and that these systems would yield roughly the same classifications. However, at lower taxonomic ranks this may no longer follow, and phenetic classifications may differ when based on sets of characters responding to differing environmental challenges. In a quantitative study of two cyclomorphic adult forms, Sokal and Thomas (1965) found that alates and stem mothers of the aphid *Pemphigus populi-transversus* share only one common factor out of five contributing to interlocality covariation (among local populations within a portion of the range of the species). Thus these organisms show little congruence. Larvae and adults of the same species will also frequently not cluster congruently.

The reasons for the lack of congruence may be twofold. Certain genes appear to be active only at specific times in the life history of an organism (Beerman and Clever, 1964). The lack of congruence may also reflect different adaptational patterns and evolutionary rates for those genes active during the development of a particular life history stage.

Factor analysis (for a detailed account, see Section 5.6) has important implications for choice of characters. Factor analysis of a matrix of correlations among a

suite of characters aims to extract the k common factors affecting the suite. Thus the covariational pattern of n characters can be expressed in terms of k factors, where $n > k$. It might therefore be argued that rather than choose n characters for an estimate of phenetic similarity, phenetic similarity should be based upon the k factors, using such characters as are necessary to evaluate the factor endowment of each OTU. There are two problems with this approach, one practical, the other theoretical. The practical problem is that in order to carry out a factor analysis one needs to have scores and correlations of the characters. Factor analysis therefore cannot really save effort, because one must record and analyze the characters before a factor analysis. Similar considerations also apply to finding clusters of highly correlated characters (which also have a high information content in a special sense; see Bisby, 1970b) with the aim of using only one character from each cluster in a subsequent analysis, and this approach carries additional dangers not possessed by factor analysis (Sneath, 1967b). Thus, unless there are theoretical advantages in estimating phenetic similarity from factor scores, it might be best to stick to the whole set of characters.

Theoretically, factor analysis seems attractive inasmuch as it reduces the number of phenetic dimensions necessary for the visualization of the organisms. But if we found that from a set of n (say 100) characters, k (say five) important factors have been extracted, should these factors be equally weighted in measuring the similarity between OTU's? There is no simple answer to this question. Should factors be weighted in terms of the number of characters they affect? This would be very difficult to evaluate. Should they be weighted in terms of the amount of variation they engender? This might be measured by the sum of squared factor loadings. Or should each factor be equally weighted regardless of its "strength" in affecting the characters? The first two approaches might yield estimates of similarity more or less approaching the intuitive appreciation of resemblances by conventional taxonomists. The third solution might yield estimates quite different from the intuitive measures, but possibly closer to the "refined similarity" produced by systematists who carefully analyze their impressions and discount the overwhelming effects of one or the other factor, such as general size, hairiness, calcification, or similar general adaptational trends. We are not prepared to say that the best estimate of similarity is the one that most closely approximates the concept as interpreted by conventional taxonomists. Rather, it should be justifiable on its own terms. More research is needed to develop an independent criterion of weighting factors.

Finally, it may be possible to use factor analysis economically to determine whether more characters should be measured. Techniques could be developed that would test additional characters to find out whether they belong to factors already known and measured or whether these new characters represent factors that have not as yet been evaluated and employed for estimating overall similarities. Sokal and Rohlf (1970) have made some experiments on these lines, and one of

TABLE 3-1

Results of studies of congruence between different sets of characters

Organisms	Comparison		Similarity coefficient	Correlation coefficient between		Source of data	
	Sets of characters	Number of characters in the respective sets		Listed comparison	Randomized sets of characters		
<i>Different body parts or systems</i>	Species of <i>Hoplitis</i> complex (bees)	Head vs. nonhead	60:62	<i>r</i>	0.61	0.81*	Michener and Sokal (1966), and Rohlf (1965)
				<i>d</i>	0.33		
	Genera of butterflies	Head vs. thorax	38:105	<i>r</i>	0.79		Ehrlich and Ehrlich (1967), and Rohlf (1965)
				<i>d</i>	0.68		
		Head vs. abdomen	38:27	<i>r</i>	0.37		
				<i>d</i>	0.28		
Thorax vs. abdomen		105:27	<i>r</i>	0.35			
	<i>d</i>		0.37				
External vs. internal	100:96	<i>r</i>	0.69	0.79*			
		<i>d</i>	0.76	0.78*			
External vs. musculature	100:75	<i>r</i>	0.70				
		<i>d</i>	0.76				
Species in two drosophilid genera	Head vs. thorax	39:51	<i>d</i>	0.45		Bächli (1971)	
	Head vs. abdomen	39:62	<i>d</i>	0.25			
	Thorax vs. abdomen	51:62	<i>d</i>	0.30			

Species of mite genera <i>Dermanyssus</i> and <i>Liponyssoides</i>	Gnathosome vs. nongnathosome	15:120	<i>r</i> <i>d</i>	0.29 0.82	Moss (1968b)	
	Dorsal idiosomal vs. nondorsal idiosomal	45:90	<i>r</i> <i>d</i>	0.35 0.80		
	Ventral idiosomal vs. nonventral idiosomal	40:95	<i>r</i> <i>d</i>	0.52 0.85		
	Leg vs. nonleg	35:100	<i>r</i> <i>d</i>	0.53 0.71		
Species of mosquito genera <i>Aedes</i> and <i>Psorophora</i>	Leg vs. nonleg	67:91	<i>r</i> <i>d</i>	0.42 0.29	Hendrickson and Sokal (1968)	
Genera of gallinaceous birds	Wing muscles vs. leg muscles	82:93	<i>r</i> difference scores assigned by investigators	0.62 0.67	P. H. A. Sneath (unpublished), Sokal and Sneath (1963) based on Hudson, Lanzillotti, and Edwards (1959), and on Hudson et al. (1966)	
	Leg muscles vs. foot muscles	77:56	S_{SM}	0.72		0.83†
	Extensor muscles vs. flexor muscles	52:81	S_{SM}	0.79		0.82†
Genera and species of Lari (birds)	External vs. skeletal	72:51	<i>r</i> <i>d</i>	0.73 0.43	Schnell (1970a,b)	
Species of angiosperm genus <i>Sarcostemma</i>	Floral vs. vegetative	61:32	<i>r</i> <i>d</i>	0.17 0.23	Johnson and Holm (1968)	

(continued)

TABLE 3-1 (continued)

Organisms	Comparison		Similarity coefficient	Correlation coefficient between		Source of data	
	Sets of characters	Number of characters in the respective sets		Listed comparison	Randomized sets of characters		
Species of angiosperm genus <i>Salix</i>	Vegetative vs. sexual	72:125	$1 - d$	0.29	0.86†	Crovello (1969)	
Strains of bacterial genus <i>Chromobacterium</i>	Morphological vs. physiological and chemical	29:76	S_j	0.61	0.75	Sneath (1972)	
<i>Different life stages</i>	Species of <i>Hoplitis</i> complex (bees)	Males vs. females	53:69	r	0.71	0.81*	Michener and Sokal (1966), and Rohlf (1965)
				d	0.35	0.67*	
Species of mosquito genus <i>Aedes</i>	Adults vs. larvae	77:71	r	0.29	0.60*	Rohlf (1963a, 1965)	
				d	0.59		0.75*
Species of angiosperm genus <i>Salix</i>	Male sexual vs. female sexual	60:65	$1 - d$	0.80	0.82†	Crovello (1969)	

NOTE: The table shows values for the matrix correlation coefficient r_{S_1, S_2} between the similarity values in matrices S_1 and S_2 derived from pairs of character sets. Different types of similarity or dissimilarity coefficients were employed. The r_{S_1, S_2} value is dependent on the numbers of characters; therefore its approximate value for random sets of characters is shown where published (or, if the value is marked with a dagger, it has been estimated by interpolation from values for slightly different numbers of characters). This dependence is revealed in the tendency for comparisons involving small numbers of characters to give low correlations; the difference between two values of r_{S_1, S_2} is a better indication of the magnitude of incongruence than is the value of r_{S_1, S_2} for the listed comparison alone. One would not expect higher correlations than those found for the randomized sets.

*Means computed from Table 2 in Rohlf (1965).

†Values estimated from the assumption that the value of $1 - r^2$ from randomized sets of n_1 and n_2 characters is approximately the square root of the sum of components proportional to $1/n_1$ and $1/n_2$ respectively.

their findings was that of three persons who chose a set of characters, one had made a selection that left three important factors unrepresented (see also Section 3.8).

In view of the partial lack of congruence it is important that characters should be distributed as widely and evenly as possible over the organisms studied. Colless (1969c) advocates a deliberate policy of stratified sampling from various organs to ensure this. He believes (personal communication) that in this way a more "switched-on" genome is sampled and that it is possible to minimize the functional, coadapted correlation in this manner, increasing the content of independent information. A somewhat different issue is whether to use published data from standard monographs. Crovello (1968g) believes that they are more reliable than is usually thought, but in common with most taxonomists, we would view such data with suspicion unless they had been collected with great attention to their comparableness, accuracy, and completeness. We draw the reader's attention to a recent review of the subject by Farris (1971).

3.7 INADMISSIBLE CHARACTERS

The proper selection of characters is clearly a critical point in the application of numerical taxonomy, as it is in other taxonomies, and misunderstandings have arisen on this score. There are, however, certain kinds of characters whose nature clearly disqualifies them from employment in a numerical taxonomic study. These are listed in the present section as inadmissible characters.

Meaningless Characters. It is undesirable to use attributes that are not a reflection of the inherent nature of the organisms themselves. For this reason taxonomists do not include the names or numbers given to specimens, nor do they employ characters whose response to the environment is so variable that it is not possible to decide what is environmentally and what is genetically determined. The number of leaves on a branch of a tree may be an example of the latter, though if acceptable evidence is forthcoming that this number is relatively constant in a species it might be admissible. However, characters affected strongly by environmental influences would be appropriate if numerical analysis were to be made of the effect of environment on phenetic relationship (Section 7.4). Similar cases may occur where other special kinds of investigation are planned. This is a matter of scientific judgment, not simply of taxonomic method, and each case must be treated on its merits.

Logically Correlated Characters. We must exclude as redundant any property that is a logical consequence of another. We cannot use both presence of hemoglobin and redness of blood if the latter is defined as possession of hemoglobin. Mathematical manipulations that constitute logical consequences should be

avoided: for example, we could not employ both the length and half the length of an organ, or the radius and the circumference of a circular structure. Similarly, characters that are tautological—those that are true by definition as well as those that are based on properties known to be obligatory—should not be included. An example of tautology is to score both tallness and height of a man. An example of a character that is true by definition is to score “presence of calcium in raphides” after having scored “raphides composed of calcium oxalate,” a substance that by definition contains calcium. To score in this instance “raphides insoluble in acetic acid” would also be scoring a character that reveals a known and invariable property of calcium oxalate, though it is not part of the definition of that compound; if this is known, the property must be omitted. In making these qualifications for admissibility we are fully aware that many or most of the “inadmissible” characters would be inadmissible on more than one count. Thus, if we use two tautological characters, we would find on examining our data after they had been prepared for machine computation that the two characters are perfectly correlated. According to our rules on empirical correlations (see below) there would be much suspicion about using both of these characters. It is quite likely that we would therefore reject one of them by the empirical correlation criterion.

Partial Logical Correlations. Many cases will arise where the dependence of one character upon another is not total but only partial. Cain and Harrison (1958, p. 89) illustrate this by an example:

Degree of melanization of the skin in mg/sq cm must not be used together with skin colour estimated by some colorimetric method if the melanin is making a contribution, which is some function of its own density, to the skin colour, unless this contribution can be subtracted from the measurement of skin colour.

We would recommend the following procedure in cases of partial logical correlations. When a character 2 depends in part upon another character 1, the decision whether to employ 2 as well as 1 will depend on the nature of the factors other than 1 that affect 2. If, to the best of our knowledge, these factors reflect heritable variations, we would include 2. But if these factors represent experimental or technical error or are otherwise unaccountable, we would not use character 2. If it is possible to partition the variation into heritable and other components it may be justifiable to use the heritable component as a character. Hall (1969b) suggests a method for reducing correlation of this kind if some estimate of this can be made. If, for example, a series of n' measurements are made of leaf widths at different places along the leaf it is obvious that these n' characters are partially correlated. Hall suggests that the contribution of character i to a resemblance coefficient should be multiplied by a weight w_i , which equals $[1 + h(n' - 1)]/n'$. Here h is an estimate of “homological indistinctness” (of the independence of the characters), varying from 0 (completely correlated) to 1 (quite independent).

Invariant Characters. We would exclude characters that are invariant over the entire sample of OTU's. To include them would not add any information about resemblances among the OTU's. Employment of invariant characters would either not affect similarity coefficients, or else would induce simple transformations of them. However, there may be instances, notably in bacteriology, where established techniques prescribe a list of tests to be performed, which thus yields invariant characters, and with protein sequences the sequence comprises the entire character set. It may be argued that from the point of view of obtaining standardized results, all characters should be included in the computation; however, since all coefficients obtained are only relative quantities, we would recommend that for most work invariant characters not be included.

Empirical Correlations. How should we decide if two characters not logically related, but highly correlated empirically, are to be counted as separate unit characters? It is possible to give extreme examples that are absurd. Thus it is observationally true to say that certain avian characters are invariably associated, and likewise certain mammalian characters. Should we attribute this effect to a single character in which birds and mammals differ—a gene, perhaps, which if mutated would turn a bird into a mammal at one jump? Clearly, we would here prefer to postulate many independent genes, and we would treat these features as independent despite the strictest correlation. In still other instances we would not assume independence so easily. The close correlation between white skin and pink eyes of total albinos in most vertebrates would be counted as a single character, since the total absence of pigment implies lack of retinal pigment.

Yet it remains true that we often need to postulate independent characters even in cases such as the albinos, for occasional albinos do have some retinal pigment. The same is true of most other apparently dependent associations. Any exception will suffice to prove that more than one character is involved. Even strictly functional associations are not as dependent as they seem at first sight: the need for the teeth to meet is only true for a species as a whole; aberrant individuals can and do occur. The fact that selection keeps two characters (the position of the top teeth and the position of the bottom teeth) in close correspondence does not necessarily imply unitary causation of these characters. In coding such a species for analysis we would employ two characters in spite of their stringent empirical correlation in the material at hand.

In serially homologous structures such as segments of an annelid or appendages of an arthropod, or in generally homologous structures such as hairs on the body surface, a character affecting equally all the members of the series could be subdivided into separate characters for each member. However, no new information would be brought about by such a procedure. In such a case we would employ only one character.

In summary, when we have evidence that more than one factor affects two correlated characters within a study, regardless of whether this evidence comes

from within the study or from outside, we would include both characters; otherwise we would employ only one. Our position is that we assume at least some independent sources of variation in any empirical correlation, unless we have reason to believe otherwise. This would err in the direction of redundancy, but it would be counterbalanced by the likelihood of obtaining new information.

3.8 THE REQUISITE NUMBER OF CHARACTERS

An important question in any numerical taxonomic study is what number of characters is required to obtain stable classifications. In spite of the large amount of work in numerical taxonomy, we still are unable to provide generally valid answers to this question. An early recommendation that no less than 60 characters should be used, whenever possible, and that if at all feasible, considerably more characters should be employed still seems reasonable, although as shall be seen, we cannot justify this requirement on either empirical or theoretical grounds. What is the correct number of characters is related to the problem of the congruence of classifications based on sets of characters from different body parts or life history stages (discussion in Section 3.6). Had the hypothesis of nonspecificity been fully valid, any set of characters would lead to sample estimates of a parametric similarity value and the question of number of characters would simply be a statistical one: what is the requisite number of characters to be sampled to obtain an estimate of similarity with confidence bands of desired width and at a desired probability level? But since it appears obvious that different sets of characters will yield somewhat different phenetic information, we cannot dispose of the problem so simply.

An early consideration has been shown to be of little utility. This was the hypothesis of the factor asymptote in which certain assumptions were made about the number of genetic factors underlying the expression of a taxonomic character and estimates were made of the number of genetic factors that could be sampled with a sample of taxonomic characters. On any reasonable assumption it was found that relatively little of the genotype would be sampled, even if a substantial number of characters were recorded. The details of the argument are furnished in *Principles of Numerical Taxonomy*. Since we are now more concerned with the phenome rather than the genome (see Section 3.6), the hypothesis has lost much of its theoretical relevance. In view of our discussion in Section 3.6 of phenetic factors as these relate to congruence, the hypothesis of the factor asymptote might be resuscitated in a new guise. Is there an asymptote for the number of phenetic factors discovered as one increases the number of characters in a study, and especially as one proceeds from one class of characters to another? Attempts to answer this question are fraught with considerable experimental and computational difficulties, but sooner or later we should be obtaining evidence on this point.

We earlier postulated the matches asymptote, which assumed that as the number of characters sampled increases, the value of the similarity coefficient becomes more

stable. This was thought to be so because one can express the resemblance between OTU's as a proportion of characters agreeing (matching), out of the total number being compared, and assume that the similarity between the two operational taxonomic units is an estimate of the parametric proportion of character matches. Such a single, definite proportion of matches (if we were able to sample all the characters) might, for example, be the matches in the nucleotide sequence of the DNA of these OTU's. If such an hypothesis were tenable, then the requisite number of characters could be simply computed from ordinary sampling theory in statistics and would need no special defense. However, it seems doubtful to us at this time that there is a single parametric measure of similarity between OTU's. It might be argued that a measure of the parametric similarity between the DNA's of the OTU's or between protein sequences could be developed. In fact, the coefficient of cross-association (sliding matches, see Section 4.4) attempts to do just that. Clearly, difficulties will be encountered by problems of nonhomologous regions, duplications, deletions, and the like. But at least theoretically we can conceive of such a measure. However, in view of our considerations about the complexity of the interactions that lead from the genetic code to the manifestations of the phenome, we doubt whether the parametric similarity between the DNA (even if it can be estimated) is a useful measure of the similarity of the organisms at the various levels of complexity we may wish to study. Yet, if phenetic similarity is not a single quantity but a shifting concept depending on the method of measurement as well as the character base, is there any hope of arriving even at an educated guess for the requisite number of characters?

Fortunately, an empirical fact may help us here. When large numbers of characters are measured, the estimate of similarity obeys what might be described as a principle of inertia. As more and more characters are added, it takes an increasingly large number of characters with quite different phenetic information to alter appreciably a given estimate of phenetic similarity. Thus, while classifications of the same OTU's based on different sets of characters might start out as different constellations in phenetic hyperspace, they would eventually converge toward the same general region, though they would not necessarily be identical as more characters were being added to the system.

The problem with an experimental test of any of this work is that it is quite difficult to obtain randomly chosen additional sets of characters. This is certainly true when the same investigator wishes to obtain additional sets of characters. His past experience in obtaining the first set of characters would undoubtedly bias him either toward avoiding characters of the sort that he had previously recorded or, on the contrary, might influence him into finding more characters like them. One way to carry out such an experiment without bias would be to have a second investigator find a number $n' > n$ characters in the hope that at least $n' - n$ new characters would be described (and hopefully more, since it is unlikely that the same n characters would be found by the second investigator). An added difficulty

is that different investigators will define identical and similar characters in different ways, and then the use of some analytical technique is required to define common "pure" characters in the two character matrices produced by the two investigators. This is a task that factor analysis might usefully perform. In the only such study to date, Sokal and Rohlf (1970) found that of 17 common factors in a study of 25 species of bees of the *Hoplitis* complex, 16 had been found by the experienced taxonomist who first classified them on the basis of 119 characters. The "least traditional" taxonomist found 14 common factors based on 53 characters, while a third inexperienced person found all 17 common factors based on 62 characters. Thus it is obvious that the results would have been largely the same regardless of the order in which the characters had been added. If the addition of further characters followed the pattern illustrated here, we would not expect the taxonomic structure to change appreciably. An interesting verification of the asymptotic approach to taxonomic stability as the number of characters increases is shown in an analysis of human populations (Jardine, 1971).

The problem could be investigated by a simulation experiment. A variable number k of character classes could be generated each containing n_i characters ($i = 1, 2, \dots, k$; and n_1, n_2, \dots, n_k not necessarily equal). Each suite of n_i characters could have its own distinct correlational pattern, with some of the characters of each class i being correlated with characters in the other classes. The distribution by which this "spillover" into other character suites could be governed might be described by two distinct models, one a J-shaped curve similar to a Poisson distribution with a low mean, the other a uniform distribution. One could then investigate to what degree the classification of t OTU's changes as one moves from one suite of characters to another. Thus, one might try the first suite of characters n_1 , compute a similarity matrix among OTU's, and then compute a similarity matrix based on the n_2 characters, as well as one on the $n_1 + n_2$ characters. Will the matrix correlation of successive similarity matrices due to additional character sets become stabilized and very high? What would be the effect of the number of characters and the distribution of characters over the character sets?

Until we have answers to these questions, and also indications of whether this model has relevance to the actual conditions occurring in nature, an answer to the question of how many characters to use is quite difficult. Caution would dictate that we consider the similarity to be open ended, hoping that its inertia will stabilize overall similarity values in a bounded region of phenetic space. The practical advice that can be given at this time is to take as many characters as is feasible and distribute them as widely as is possible over the various body regions, life history stages, tissues, and levels of organization of the organisms. Since congruence is always less than is expected from random samples of characters, the number of characters used will set a lower limit to the confidence levels of the similarity coefficients. The investigator should therefore employ at the very least as many characters as will give the confidence limits he wishes (see Section 4.10). Since we believe that there

will be correlation between suites of successively taken characters and that the phenomenon of inertia will exert a stabilizing effect on similarities, we do not feel as strongly as Ehrlich and Ehrlich (1967) that overall similarity cannot be dealt with in practice.

3.9 THE PROBLEM OF CHARACTER WEIGHTING

The problem of weighting characters is apt to give rise to misunderstanding, and we have therefore attempted to clarify this. Numerical taxonomists are generally in agreement in giving each character equal weight when creating taxonomic groups, although Burt (1964) has pointed out that numerical taxonomy need not necessarily be based on equally weighted characters: he calls such equally weighted classifications *isocratic*. We ourselves believe that equal weighting is desirable, and discuss here its justification.

We should first emphasize that we are not here discussing the use of characters in identification. After a manner of speaking, "weighting" is used in such a procedure and properly so. However, the construction of taxonomic keys and the identification of specimens belong to a later stage of taxonomic procedure subsequent to the formation of the taxa concerned. We are therefore discussing *a priori weighting*, before a classification is commenced, and what we feel is objectionable is to presuppose knowledge that is not yet available, either about the classification of the organisms, or about the presumed significance of their characters.

We also emphasize that we are not advocating equal weight for character complexes, such as flowers or leaves. These are broken down into their unit characters, and hence effectively receive weights in proportion to their complexity or information content. Unit characters should quite appropriately receive unit weight. Expressed thus, the problem does not appear so illogical.

Kendrick (1964, 1965) has suggested that characters should be weighted depending on the "rank" of the organs which they describe. Thus, for example, if two OTU's **a** and **b** have leaves with several "subcharacters" such as "smooth-hairy, simple-compound," etc., while OTU **c** has no leaves, a situation could arise wherein **a** and **b** would differ in the entire set of subcharacters, agreeing only in the character "possession of leaves." By contrast either **a** or **b** could be compared with **c** only in the character "possession of leaves" for which they would differ. Thus the similarity between **a** and **b** with respect to leaves might be less than that between either of the two and **c**, a situation Kendrick feels is undesirable. OTU's **a** and **b** would appear to be more dissimilar than either is from **c** simply because the dissimilarities between **a** and **c** cannot be specified and the subcharacters for **c** would have to be left unscored. However, we believe that if we were to take a representative sample of characters from the entire organism, it is unlikely that we would get so distorted a picture of the dissimilarities. In fact by some methods of measuring resemblances, **a** and **b** would be no more dissimilar than either is to **c**.

To support this point of view let us take the absurdly extreme case wherein **a** and **b** are two flowering plants with different leaves and **c** is a fungus. It seems disturbing to find that **a** and **b** are more dissimilar with respect to leaf morphology than **a** and **c**, which differ simply in their possession of leaves. We are disturbed by this disproportionate dissimilarity because we recognize the fungus as being so very different from the flowering plants. But if we include the entire repertoire of morphological, histological, and physiological characteristics possessed by the angiosperm OTU's and the fungus it is obvious that the great similarity of the former to each other will emerge in contrast with their marked dissimilarity to the fungus. By taking only the characters of the leaves we are selecting a very specific fraction of the phenome. This point also relates to how one should treat inapplicable comparisons, and is a point returned to in Section 4.12.

In addition, it should also be made clear that we are not here discussing the mathematical weighting attendant on different coding and scaling procedures (Section 4.8). "Weighting" produced that way is directed to the specific purpose of the classification and the estimation of different components of phenetic resemblance, such as the removal of a general size factor or the kind of taxonomic structure we wish to evaluate. It is of course true that characters not employed are automatically given zero weight, but this is inevitable by any procedure. It does not justify arbitrary or irrational weights for those that are employed.

That every character should be given equal weight is implicit in the work of Adanson and the writings of Gilmour (1937, 1940, 1951) and Cain and Harrison (1958); and it was stated explicitly by Sneath (1957a) and Michener and Sokal (1957). Sneath reached this conclusion on considerations stemming from Gilmour's work on epistemology. Since natural taxa ideally contain the greatest possible content of implied information, this can only be measured in the number of statements that can be made about its members, which is independent of how important we may think any statement is. This argument has been developed at some length in connection with the "general" nature of natural classifications (see Sneath, 1958). Michener and Sokal (1957) concluded that even if desirable, there is no rational way of allocating weight to characters and therefore one must in practice give them all equal weight. In addition, when many characters are employed, the statistical analysis of similarity is only slightly affected by weighting some characters (unless this weighting is extreme). Thus Moss (1968b) noted that an accidental error, which weighted 61 character states of 18 characters nearly a thousandfold, nevertheless had a negligible effect on the resemblance coefficients in a study based on 135 characters and 17 OTU's. It is perhaps worth reemphasizing this point, for in practice the measures of overall similarity do yield substantially the same results when based on many characters, whatever the weighting (within reason) of individual characters, it would seem unnecessary to argue the point further. It would be useful if further studies could be made of the extent to which arbitrary weights

affect taxonomy, in order to provide some guidelines for avoiding excessive distortion.

Some other suggestions have been made for weighting. Thus Williams, Dale, and Macnaughton-Smith (1963) have proposed for ecological applications that characters be weighted in proportion to their average correlation with all other characters. This is likely to have much the same effect as the suggestion by Mayr (1964) that one should first make a classification on unweighted characters and from this obtain the character weights (presumably according to their relative value in separating the taxa found) and second, repeat the classification using these *a posteriori* weights. Jizba (1964) has tried this with geological material, but his second, third, and even fourth classifications differed little from his first. Similar experiments in biological taxonomy are needed. By such a procedure the taxa might perhaps be somewhat more sharply demarcated, but even this is in doubt. One would have to ask whether iterative computer runs using reiterated series of weights would be desirable, and what the criterion for stopping iterations would be.

It is possible to use the rarity of features, or alternatively their commonness, in the whole set of OTU's to weight the characters, but the results are likely to be little different (except perhaps by a scaling factor, e.g., the example given by Sneath, 1965). Smirnov (1968) has suggested a similar kind of weighting, which is discussed more fully in Section 4.4 though we feel it has little to commend it.

Methods of numerical cladistics weight characters or character states as a result of explicit algorithms during the clustering procedure. Investigators may or may not agree with the aptness of these algorithms but their explicitness, and the fact that weight is not a priori, distinguish this procedure from weighting as practiced in conventional taxonomy.

One such method weights characters according to some function of their variability within OTU's (generally populations or species). Thus Farris (1966) suggests that characters varying little within populations are more reliable indicators of cladistic relationship than variable characters; invariable characters could therefore be weighted more heavily for cladistic studies. Eades (1970) makes similar proposals. However, this generalization about characters and evolution seems to us to be still uncertain. More interesting is the suggestion (Farris, 1969b) that weighting of characters for cladistics could be in proportion to the degree to which they fit a parsimonious cladogram, the aim being to improve iteratively the fit and the parsimony, as discussed further in Section 6.4. Goodman (1969) has a somewhat different aim in proposing that a character difference should be weighted inversely according to the within-OTU variance, because in his studies on plant breeding the variance may be largely due to environmental effects, so that genotypic differences might be better expressed if variable characters are deemphasized. Also variance may differ markedly within different OTU's. This particular scheme has the drawback that the weight becomes indeterminately large for characters constant within OTU's, so that a few of these characters may dominate the analyses. A better

suggestion, by Flake, von Rudloff, and Turner (1969), is to multiply the squared difference in character i by a weight w_i^2 , where w_i is $1 - (s_{i, \text{within}}/s_{i, \text{total}})$, where $s_{i, \text{within}}$ is the average within-OTU standard deviation for character i , and $s_{i, \text{total}}$ is the total standard deviation. If w_i is negative it is taken to be zero, that is, the character is excluded as being too variable within OTU's for safe use.

All the methods mentioned in the last paragraph give most weight to those characters that vary least within OTU's. Such weighting increases the ease with which taxa can be discriminated, and thus leads towards generalized distance, further taken up in Sections 4.3 and 8.5. We would consider these as methods for special purposes rather than for general taxonomic analyses. It may be noted too that Crovello's character state difference (Crovello, 1968e; Section 4.13) gives less weight to characters that have the same degree of variation within all OTU's, and more to those that are highly variable within some OTU's and invariant within others. However, this effect is unlikely to be very pronounced in practice.

The arguments in favor of equal weighting fall under seven headings.

1. If it cannot be decided how to weight the features, one must give them equal weight—unless it is proposed to allocate weight on irrational grounds.
2. To create taxonomic groups, one must first decide how to weight the features that are to be employed for classifications. Therefore, one can use no criterion that presupposes the existence of these taxa. For example, one cannot choose the constant features—to know if they *are* constant one must first set up taxonomic groups, and these have not yet been established.
3. The concept of taxonomic importance has no exact meaning. If “importance” means “importance to me because I am interested in it,” this is only special pleading. If “importance” means basic or fundamental, this can only mean that it sums up a number of other characters: if they are unknown, they are hypothetical; if known, the character is not single but multiple. If “importance” means essential to survival, the taxonomy can estimate viability but not resemblance. If “importance” means “correlation with other features,” then the added weight is due to these other features; where we observe the correlation breaking down, we do not regard the feature as important.
4. If differential weighting is admitted, exact rules must be given for estimating it. One must know whether the weight to be given to the possession of feathers is twice or twenty or two hundred times that given to possession of claws, and why. We do not know of any method for estimating this, and even if such a method were to be developed we doubt if any systematist would have the patience to use it because of the hundreds of characters he would need.
5. The nature of a taxonomy depends upon its purpose: a systematist could arrange living creatures in many ways but chooses one way because he thinks it is the best for some purpose. We hold the view that a “natural” or “orthodox” taxonomy is a general arrangement intended for general use by all kinds of

scientists (Gilmour, 1937; Sneath, 1958). It cannot therefore give greater weight to features of one sort, or it ceases to be a general arrangement. Being general, it is best for general purposes but is perfect for none.

6. The property of "naturalness" is, we believe, due to the high content of implied information that is possessed by a natural group. A group such as the Mammalia at once tells us much about its members with a high degree of certainty. A group such as "black animals" tells us nothing more than that they are all black. The content of information is measured by the number of statements that can be made about its members: each statement has unit value, and whether we think them important or not is irrelevant.

7. The use of many characters greatly evens out the effective weight that each character contributes to the similarity coefficient. Unless highly unequal weights are given to some characters, the very employment of many characters tends to make the taxonomy equally weighted.

Equal weighting can therefore be defended on several independent grounds: it is the only practical solution, it and only it can give the sort of natural taxonomy that we want, and it will appear automatically during the mathematical manipulations. Singly, these arguments are cogent; taken together, we think they are overwhelming.