

Taxonomic Principles

To anyone familiar with current developments in taxonomy, it is clear that taxonomy is undergoing a period of rapid conceptual and procedural change, probably unequaled since the immediate post-Darwinian era. Numerous publications in journals such as *Systematic Zoology*, *Evolution*, and *Taxon* reflect the effervescence of ideas that have suddenly enlivened this field. The preceding period of similarly rapid progress in biological systematics has been called the "New Systematics"; it may be said to have begun with the appearance of the book of the same name edited by Huxley (1940). Advances in genetics, cytology, and geographic variation during the period led to considerable progress in the understanding of evolutionary mechanisms at the species and infraspecies levels. However, the New Systematics contributed little to our understanding of the nature and evolution of the higher categories and of taxonomic structure in general. Books such as those by Rensch (1947), Schmalhausen (1949), Simpson (1953), and Hennig (1966) deal with the latter topics but they contain little more than descriptive generalizations. The failure of the New Systematics to provide an adequate base for animal taxonomy is also discussed by Blackwelder (1967a, p. 336).

Traditional taxonomy attempts to fulfill too many functions and as a consequence fulfills none of them well. It attempts (1) to classify, (2) to name, (3) to indicate degree of resemblance, and (4) to show relationship by descent—all at the same time. We shall show in separate sections in this chapter that it is impossible not only in practice but also in theory for the traditional system to perform these

tasks adequately. In Section 2.1 we shall discuss empirical and operational approaches to the subject. Section 2.2 will concern itself with the problems of the natural system; in Section 2.3 we shall analyze taxonomic relationships. Problems of phenetic taxonomy are detailed in Section 2.4 and those of cladistic taxonomy are detailed in Section 2.5. A decision must be made on which of these methods is to serve as a basis of classification; this will be found in Section 2.6. The last two sections, 2.7 and 2.8, deal with problems of taxonomic rank and with desirable properties of a taxonomic system, respectively.

2.1 EMPIRICAL AND OPERATIONAL APPROACHES

Recent statements on taxonomic theory (Ehrlich and Holm, 1962; Davis and Heywood, 1963; Sokal and Camin, 1965; Jardine and Sibson, 1971) have stressed the *empirical approach* in taxonomic work. The main emphasis in empirical taxonomy is to base classifications and taxonomic judgment on firm observation and not upon "phylogenetic" assumptions. We think the theoretical framework of conventional phylogenetic systematics has served as a straitjacket for taxonomic concepts and ideas at all levels, from that of the infraspecific population to that of phyla; in the final analysis it is evolutionary theory that has suffered as well, since it is only as good as the taxonomic data fed into it. Thus, so long as taxonomic descriptions and judgments are made to conform to concepts such as the biological species (Mayr, 1963), or monophyly (Simpson, 1961), taxonomists tend to consider difficulties and discrepancies as embarrassing exceptions to generally accepted principles. But in recent years the emphasis has been on the description of those variational patterns that do in fact exist in nature. New principles will emerge from such studies and a body of theory whose heuristic aspects are beginning to emerge is already forming (see Chapters 6 and 7, and Section 10.3).

The emphasis on empirical analysis of taxonomic data led naturally to what has become known as the *operational approach* to taxonomy (Sokal and Camin, 1965) by analogy and extension of P. W. Bridgman's ideas. In our context operationism implies that statements and hypotheses about nature be subject to meaningful questions; that is, those that can be tested by observation and experiment. One must establish criteria for defining categories and operations; otherwise it would be impossible to engage in a meaningful scientific dialogue about them. Sokal and Camin illustrate this concept by stating that

... if we wished to determine whether A is more related to B than it is to C, we have to give clear definitions of what we mean by "more related," i.e., by what criteria more or less relatedness can be measured, and we must be able to issue a series of instructions by which we, our assistants, or our colleagues can determine the relationships and solve the problem originally posed. We must be certain that the data we work with are subject to definable logical operations. Secondly, we must be concerned that the operations carried out to answer questions raised about the material are such that they can be communicated unambiguously to other intelligent persons as well as to machines able to handle the logic and computation required.

Such a guiding principle seems natural and desirable to us and we are unaware of any substantial arguments advanced against this approach in taxonomic work. The term operationism, however, is something of a red flag to certain philosophers of science (Hawkins, 1964; Hull, 1967, 1968a). Hawkins (1964) points out that emphasis on operational procedures and strictly quantitative measurements may be unwarranted in the biological and social sciences, since the theory to which the body of data pertains is itself ill-defined because of the complexity of the subject matter: "For the purposes of... crude theory, crude measurements suffice." Hawkins believes that in biological taxonomy the conventional types of observations are sufficiently reliable to serve the needs of existing theory. We would differ on this point, since we feel that useful theory cannot be obtained in this field until more precise measurements and definable operations have become standard procedure in biological taxonomy.

Hull (1968a) points out that operationism carried to the extreme would lead to inference-free direct observations, unlikely to lead to theoretical advances. He adds that if, in contrast, empirical taxonomy itself relies on inference and theory, its proponents should not condemn phylogenetic systematists for their excursions into similar fields.

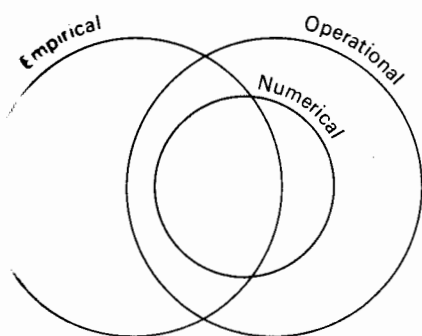
Not all concepts can be made equally operational and some not at all. Nor is it necessary, as Hull (1968a) points out, that operationism be extended to all members of a class of concepts.

We do not object to nonoperational concepts categorically, although we would always prefer more-operational concepts to less-operational ones; but when the nonoperational concepts are vague and ill-defined, and have no heuristic value, we are opposed to them. Concepts such as the biological species (*sensu* Mayr et al.) and the phylogenetic notion of homology were vital in the development of modern biological theory. Today they are more of a hindrance than a stepping stone to new discoveries and it is for this reason as much as their low operational value that we wish to redefine them or possibly even dispense with them.

Operational and empirical taxonomy (by which we mean operational procedures during classification, and empirical observations of taxonomic data) are not wholly congruent. Although much of empirical taxonomy is operational it is conceivable that nonoperational criteria could be used to process empirical observations (Figure 2-1). Numerical taxonomy as generally practiced is both operational and empirical. Although it is quite possible to visualize a numerical taxonomy that is not empirical, all numerical taxonomies are likely to be operational.

2.2 THE NATURAL SYSTEM

Great difficulties have always accompanied attempts at defining a natural system. Thus we have the significant comment of Linnaeus himself (mentioned by de Candolle, 1813, p. 60) that he had been unable to discover a natural method of



Taxonomy

FIGURE 2-1

Diagram of the relationship between operational, empirical, and numerical taxonomies. [From Sokal and Camin (1965).]

classification. Until recently, apart from the work of Gilmour mentioned later, very little was written on this. Danser (1950) realized the difficulty of defining natural groups but was not able to state any exact or scientific definition for them, ending with the hope that "... some day systematics will arrive at a more exact stage, but this does not alter the fact that already now we are entitled to face its problems, be it for the moment in a more intuitive but nevertheless scientific manner." Simpson (1961, p. 57) agrees that "in fact much of the theoretical discussion in the history of taxonomy has, beneath its impersonal language and objective facade, been an attempt to find some theoretical basis for these personal and subjective results."

The term "natural classification" has had a variety of meanings. In the early days of systematics it meant a classification that was in accord with nature, but this meaning was undefined further (though it was commonly implied to describe the opposite of artificial or arbitrary systems, but these were not well defined themselves). Later it came to mean variously a taxonomy based on maximum correlations between characters, or one based on phylogeny, and the latter sense is one in which traditional taxonomists use it today, though we explain later why we feel this is not a suitable usage. The development of the concept of the natural system is briefly discussed in the following pages.

The earliest attempts at systematics were based, as Cain (1958) has shown, on Aristotelian logic. This was the method used by early systematists such as Cesalpino (1519-1603) and even largely by Linnaeus (1707-1778). The Aristotelian system as applied to taxonomy consisted in the attempt to discover and define the *essence* of a taxonomic group (what we may somewhat loosely think of as its "real nature" or "what makes the thing what it is"). In Aristotelian logic this essence is expressed in axioms that give rise to properties that are inevitable consequences; for example, the essence of a triangle on a plane surface is expressed by its definition as a figure bounded by three straight sides, and an inevitable consequence is that any two sides together are longer than the third. Such logical systems are known as systems

of *analyzed entities*, and early systematists supposed that biological classifications could be of this kind. The terms *genus* and *species* had technical meanings in logic, and these were taken over into taxonomy. These points are well discussed by Thompson (1952) and Cain (1958, 1962). Aristotelian logic does not, however, lend itself to biological taxonomy, which is a system of *unanalyzed entities*, whose properties cannot be inferred from the definitions—at least not if the taxonomy is to be a natural one in any of the usual senses.

Caspar Bauhin (1550–1624), John Ray (1627?–1705), and Pierre Magnol (1638–1715) had a strong intuitive sense of what natural taxa were, although they did not express themselves clearly. This is what de Candolle (1813, p. 66) aptly called “groping” (*tâtonnement*). According to de Candolle, Magnol claimed to have a clear idea of a natural family of plants even though he could not point to any one character which was diagnostic of that family.

This comment by Magnol and a similar comment by Ray (quoted by Cain, 1959c) were among the first admissions that it might not be possible to find *any single* diagnostic character for a natural taxonomic group. This is a point of the very greatest importance, which can scarcely be overemphasized. Michener (1970) supports this view for all taxa but species; others, among them Sneath (1957a) and Jardine and Sibson (1971), state that it is true of natural taxa of *any* rank. While “artificial” or “arbitrary” taxa can indeed be defined by a single character, this is not necessarily true of natural taxa. Every systematist knows of instances where a character previously considered to be diagnostic of a taxon is lacking in a newly discovered organism that clearly belongs to the taxon. A striking example is the lack in some species of fish of red blood corpuscles (Ruud, 1954), hitherto considered to be an invariable attribute of all vertebrates. Fortunately, as Michener (1957) says, natural taxa generally do possess some distinctive characters in practice although they need not do so in theory.

Biologists are indebted to Beckner (1959) for a clear enunciation of the important concept of “polytypic” natural taxa. Its implications in taxonomy have been discussed by Simpson (1961, pp. 41–57). Since this term and its converse, “monotypic,” have meanings already well established in systematics, the substitute terms “polythetic” and “monothetic” suggested by Sneath (1962) have come into general use. We have assumed, in interpreting Beckner’s concepts, that he intended the term “property” to be taken in the general sense of some definite value of a taxonomic character.

The idea ruling *monothetic groups* is that they are formed by rigid and successive logical divisions so that the possession of a unique set of features is both sufficient and necessary for membership in the group thus defined. They are called monothetic because the defining set of features is unique. That is, all the members of any group possess all of the features that are used to define that group. Any monothetic system (such as that of Maccacaro, 1958, or that of Williams and Lambert, 1959) will always carry the risk of serious misclassification if we wish to make natural

phenetic groups. This is because an organism that happens to be aberrant in the character state (see Section 3.2) used to make the primary division will inevitably be moved to a taxon away from the required position, even if it is identical with its natural congeners in every other character state. The disadvantage of monothetic groups is that they do not yield “natural” taxa, except by a lucky choice of the character used for division. The advantage of monothetic groups is that keys and hierarchies are readily made.

By contrast, in a *polythetic group*, organisms are placed together that have the greatest number of shared character states, and no single state is either essential to group membership or is sufficient to make an organism a member of the group. This concept was stated many years ago (for example, by Jevons, 1877, pp. 682–698) and more recently by Kaplan and Schott (1951) and Gasking (1960). For its formal expression we cannot do better than to quote Beckner’s definition (1959, p. 22):

A class is ordinarily defined by reference to a set of properties which are both necessary and sufficient (by stipulation) for membership in the class. It is possible, however, to define a group K in terms of a set G of properties f_1, f_2, \dots, f_n in a different manner. Suppose we have an aggregation of individuals (we shall not as yet call them a class) such that :

1. Each one possesses a large (but unspecified) number of the properties in G .
2. Each f in G is possessed by large numbers of these individuals and
3. No f in G is possessed by every individual in the aggregate.

By the terms of 3, no f is necessary for membership in this aggregate; and nothing has been said to either warrant or rule out the possibility that some f in G is sufficient for membership in the aggregate.

He then goes on to say that a class is polythetic if the first two conditions are fulfilled and is fully polythetic if condition 3 is also fulfilled. He points out that taxonomic groups are polythetic classes, but that polythetic concepts are by no means restricted to taxonomy or even to biology, for Wittgenstein emphasized their importance in ordinary language and especially in philosophy—polythetic ideas are implied by the concepts of “meaning,” “referring,” “description,” and so on. There is a close parallel between Wittgenstein’s “family resemblance” and taxonomic resemblance. As we have noted above, natural taxa are usually not fully polythetic, since one can usually find some character states common to all members of a taxon. It is possible that they are never fully polythetic, because there may be some character states (or alleles) that are identical in all members of a given taxon; even if there are many alleles or pseudoalleles of a gene, there may well be parts of the gene that are identical in all members. Recent work on protein sequences (discussed at greater length in Section 3.5) suggests that there are at least parts of genes that are most probably invariant in all members of a taxon, while work in population genetics (for example, that of Kimura, 1968) implies that the number of alleles at a locus in a population is limited by the mechanism of gene replacement during evolution. Both ideas convey the implication that taxa are not fully

polythetic. Nevertheless, for practical purposes we must consider the possibility of a taxon being fully polythetic, since we cannot be sure that we have observed any characters that are common to all members.

Beckner also points out the importance of condition 2. If, for example, the various *f*'s are found in only one individual of the aggregate, then each individual will possess a unique subset of the *f*'s and will share no *f*'s with any other individual. Such a situation does not yield a polythetic class. This is well discussed by Hull (1965). Let us illustrate such cases with the aid of two-state (presence-absence) characters: for example, individuals **a**, **b**, **c**, and **d**, do not form a polythetic class with the respective *f*'s (presence or plus-states) of characters {1, 2, 3}; {4, 5, 6}; {7, 8, 9}; and {10, 11, 12}. If, however, as in the table shown below, individual **a** possesses {1, 2, 3}; individual **b** possesses {2, 3, 4, 5}; individual **c** possesses {1, 2, 4, 6}; and individual **d** possesses {1, 3, 4}; then the class of {**a**, **b**, **c**, **d**} is polythetic (and in this instance is also fully polythetic, since no one character state is found in all the four individuals). This may be displayed in an arrangement such as this one:

Characters	Individuals					
	a	b	c	d	e	f
1	+	-	+	+	-	-
2	+	+	+	-	-	-
3	+	+	-	+	-	-
4	-	+	+	+	-	-
5	-	+	-	-	+	+
6	-	-	+	-	+	+

Individuals **e** and **f**, however, form a fully monothetic group.

One of the difficulties of Beckner's definition is that in natural taxa we do commonly have *f*'s that are not possessed by large numbers of the class. Furthermore, we cannot test whether any given *f* is possessed by large numbers of the class before we have made the class, and therefore we cannot decide whether to admit this *f* into the set *G*. This difficulty can be avoided by defining class membership in terms of common (or shared) attributes. Polythetic groups can of course themselves be arranged polythetically to give higher polythetic groups, as is done in building a hierarchy in the natural system. The advantages of polythetic groups are that they are "natural," have a high content of information, and are useful for many purposes. Their disadvantages are that they may partly overlap one another (so that hierarchies and keys are less easy to make than with monothetic groups) and that they are not perfectly suited for any single purpose.

An important practical difference between "classification from below" (agglomerative clustering—the grouping of individuals into species, species into genera,

genera into tribes, tribes into families, and so on) and “classification from above” (divisive clustering—the division of the kingdoms into phyla, phyla into classes, and so on) is that the latter process is usually based on monothetic criteria. Classification from above therefore carries the risk that the divisions do not give “natural” taxa, yet it is a necessary practice in order to isolate a group of organisms of a manageable size for study. The important point is that the group classified from above may be incomplete or very heterogeneous; that is, some of its closest relatives may have been omitted, either through ignorance or because the forms have been misclassified.

A thorough early reevaluation of systematics was made by Adanson, a botanist of independent and original views. He rejected the a priori assumptions on the importance of different characters (which were a consequence of Aristotelian logic); he correctly realized that natural taxa are based on the concept of “similarity”—which is measured by taking all characters into consideration—and that the taxa are separated from each other by means of correlated features (Adanson, 1763, pp. clv, clxiv). The method he used was very cumbersome. He made a number of separate classifications, each based on one character, and examined them to find which classifications divided up the creatures in the same way. These classifications he took as indicating the most natural divisions, which were, of course, therefore based on the maximum correlations among the characters. By treating every character in the same way he was in effect giving them equal weight; it was upon this important corollary that his contemporaries attacked him (see de Candolle, 1813, pp. 70–72), without realizing that their own beliefs on the relative importance of various characters, far from being based on a priori assumptions as they imagined, were in fact a posteriori deductions from intuitive taxonomies of precisely the kind Adanson was recommending (Sneath, 1957a; Cain, 1959a,b). Adanson’s earliest work in this direction was on molluscs (1757). No other workers, except perhaps Vicq-d’Azyr (1792) and Whewell (1840), seem to have followed up Adanson’s ideas until recently.

We may ask why Adanson’s method, though excellent in theory, was a failure in practice. Stearn (1961) considers that the material available in Adanson’s day was too limited to allow of success, and we may add that such methods were quite impracticable before the advent of computing machines. Nevertheless, as de Candolle admitted, Adanson’s taxa were for the most part more natural than earlier arrangements, although the superiority was not very marked. A fuller review of Adanson’s contributions is given in Sokal and Sneath (1963) and by Stafleu (1963) and Sneath (1964c); for some other views see Jacobs (1966), Leroy (1967), Guédès (1967), and Burt (1966). Burt charges that the numerical taxonomists misinterpret Adanson’s ideas and more specifically that Sneath (1964c) is wrong in proclaiming Adanson to be the father of numerical taxonomy. We prefer to let historians of science pursue this argument. For although it was—and remains—important to trace the roots of the historical origins of an idea in science, the

development of numerical taxonomy has so far outpaced the early primitive ideas on this subject that to have to rely on Adanson's views for a validation of modern numerical phenetics seems as irrelevant as to rely on Mendel's writings for a validation of the findings of the molecular geneticists.

In the preevolutionary days of systematics it had been found empirically that a nested, hierarchical system gave the most satisfactory and "natural" arrangement of the data. Such a system could generally be constructed on the basis of a few characters. The art of the practice lay in finding suitable characters, to prevent the classification from creating strange bedfellows clearly incongruous when judged by their great differences in other characteristics. There was little attempt either to understand why it was possible to construct a system or to discover the rational method of choosing the "right" characters. We discuss below the development of the understanding of what it is that makes taxonomic groups "natural" and how it is possible *after creating such natural taxa* to discover characters that are suitable for discriminating between them.

Until the impact of the theory of evolution, the subsequent development of systematics took place largely in France (de Candolle, A. L. de Jussieu, Cuvier, and Lamarck) and was in the direction of greater sophistication on the theme of the coordination of characters into a harmonious whole. This was carried even to the point of implying that a whole animal could be reconstructed from one bone. One can, of course, identify a known animal from one bone, but to reconstruct from it a new animal with all its soft parts is a feat of a different order, as Simpson (1961, p. 44) points out.

The advent of the theory of evolution changed the practice of systematics very little, although the professed philosophical basis of systematics was radically altered. Natural classifications were considered to be those established on the basis of monophyletic taxa (see also Section 2.3). The theory of evolution did, however, provide a credible explanation for the nested distribution of taxa.

Little more was done on taxonomic theory until the conceptual basis of natural taxonomies was discussed from the standpoint of logic in a classic paper by Gilmour (1937) and expanded in later works (Gilmour, 1940, 1951, 1961; Gilmour and Walters, 1963). He pointed out that logicians have long realized that the central idea underlying "natural" groupings is the great usefulness of a method that can group together entities in such a way that members of a group possess many attributes in common. Indeed, we maintain that the elusive property of naturalness is simply the degree to which this principle obtains. The idea of overall similarity follows from this and is a function of the individual similarities in each of the many characters over which two entities are being compared. As Gilmour points out, natural classifications are not restricted to biological ones (see Chapter 11).

One of Gilmour's main points is that the nature of a taxonomy depends on its purpose. We could arrange living creatures in many ways, but we choose one way because we think it is best for some purpose. If the purpose is restricted, then the

classification is a special classification, often called "arbitrary." Such a classification conveys less information than a general or "natural" one. For example, we can divide mammals into carnivores and herbivores for the purpose of ecology; then the designation "carnivore" only tells us the kind of food they eat. We hold the view with Gilmour that a "natural" taxonomy is a general arrangement intended for general use by all scientists. In addition, intermediate situations can occur between the highly natural (such as the class Mammalia) and the wholly artificial (such as creatures whose generic names begin with the letter "A"). An example of a partly natural group is the group that gardeners call "Alpines"—plants that share numerous growth and physiological characteristics reflecting their adaptation to alpine conditions. Edwards and Cavalli-Sforza (1964) have pointed out that general or natural classifications are not too tightly defined, but the terms have nevertheless considerable value in illustrating the basic logic that underlies taxonomy. Mayr (1969a, p. 67) states that "the concept [of a natural system] is so permeated with essentialist-creationist ideology that its use invariably evokes a misconception among nontaxonomists." Since he does not believe in the existence of the "Natural System," he prefers not to use the term at all.

We believe that natural classifications are of great usefulness because when the members of a group share many correlated attributes, the "implied information" or "content of information" (Sneath, 1957a) is high; this amounts to Gilmour's dictum that a system of classification is the more natural the more propositions can be made regarding its constituent classes. Remane (1956, p. 4) tries to show that the predictive value of taxonomic groups is only true of natural taxa, not of artificial ones. It is obvious that artificial groups established on a single character are of low predictive value. Nevertheless, such groups may by chance prove to be partly natural, since such a single character may be highly correlated with the other characters of the taxa in question. It would be possible to devise a measure of the extent to which this is true of any character in any given taxonomic system. Such techniques are discussed in some detail in connection with probabilistic similarity coefficients in Section 4.6, and with identification problems in Chapter 8, but a brief account follows here.

The concept of information content is one that may not be easy to visualize without an example. We therefore give below a very simple illustration of the principle involved. When using small numbers of specimens and characters one cannot include many "atypical" characters if the groupings are to be reasonably sharp, but a few have been included to show that the principle applies to polythetic groups as well as monothetic ones.

Suppose we have ten specimens scored for five qualitative characters, and the "natural" groups are **A** and **B** as shown below. We can now count, for each group in turn, the number of characters about which we can be reasonably certain, if we were given a randomly sampled specimen of that group. By extension we may make similar statements about a new specimen, which has been allocated with complete

certainty to the group in question. Whether a confident statement about the expected state of any one character is possible is shown in the column headed "Confident statement possible?" for both groups. The predicted state of the character is shown in parentheses.

We need to decide on some level of confidence, and purely for illustration we have used a level that is not very stringent, having an 80 percent (or better) chance of being correct. The reader may if he wishes use other levels, such as the most stringent possible (100 percent), and satisfy himself that the general behavior of the system is the same. If the level is reduced to 50 percent, the system loses its power of prediction, which is then effectively nil. The division into groups **A** and **B** is:

Characters	Group A						Confident statement possible?	Group B				Confident statement possible?
	Specimens							Specimens				
	a	b	c	d	e	f		g	h	i	j	
1	+	+	+	+	+	+	Yes (+)	-	-	-	-	Yes (-)
2	+	-	+	+	+	+	Yes (+)	-	-	-	-	Yes (-)
3	-	-	-	-	+	-	Yes (-)	+	+	+	+	Yes (+)
4	-	-	-	-	-	-	Yes (-)	+	-	+	+	No
5	-	-	-	-	-	-	Yes (-)	+	+	+	+	Yes (+)
Numbers of confident statements possible							5					4

There are five confident statements possible about group **A** but only four for group **B** (because character 4 does not have a constancy of 80 percent in group **B**), giving a total of nine statements for both groups.

Suppose we now divide the ten specimens differently, so that **e** and **f** are transferred to group **B**, yielding the new groups **A'** and **B'**. The table now becomes:

Characters	Group A'				Confident statement possible?	Group B'						Confident statement possible?
	Specimens					Specimens						
	a	b	c	d		e	f	g	h	i	j	
1	+	+	+	+	Yes (+)	+	+	-	-	-	-	No
2	+	-	+	+	No	+	+	-	-	-	-	No
3	-	-	-	-	Yes (-)	+	-	+	+	+	+	Yes (+)
4	-	-	-	-	Yes (-)	-	-	+	-	+	+	No
5	-	-	-	-	Yes (-)	-	-	+	+	+	+	No
Numbers of confident statements possible					4							1

It is now seen that the numbers of confident statements possible have decreased for both groups, and the total is now only five. In this example the transfer of only one specimen from **A** to **B** will not reduce the number of confident statements at the 80 percent level (unless specimen **e** is transferred), but it will at the 90 percent level. This kind of measure of naturalness is very close to what Gilmour was suggesting, but in practice it is more satisfactory to use other measures of information (see Section 4.6 and Chapter 8).

A natural classification can be used for a great variety of purposes, while an artificial one serves only the limited purpose for which it was constructed. As Sneath (1958) has emphasized, natural or "general" classifications can never be perfect for all purposes, since this is a consequence of the way natural groupings are made. By putting together entities with the highest proportion of shared attributes, taxonomists refrain from insisting that the taxa shall share any particular attribute, as a very simple trial would show. This is the reason for emphasizing the historical importance of the realization that natural taxa do not necessarily possess any single specified feature. This spelled the doom of the Aristotelian concept of an essence of a taxon, for natural groups are in logic unanalyzed entities. Simpson (1961) rejects as illogical the contention by Gilmour (1951) that a classification serving a large number of purposes will be more natural than one which is more specialized and that the most useful and generally applicable classification will be the most natural one. We feel that Gilmour's usage corresponds to the intuitive sense of naturalness which taxonomists have possessed since even before Darwin. This usage is now becoming more widely accepted in taxonomy (e.g., Davis and Heywood, 1963). Gilmour's dictum—that a system of classification is the more natural the more propositions can be made regarding its constituent classes—admits of objective measurement and testing, in contradistinction to Simpson's natural system. Furthermore, Gilmour's system has powerful predictive properties; it is therefore the one we recommend.

2.3 TAXONOMIC RELATIONSHIPS

There has been much recent progress in understanding the nature and kinds of taxonomic relationships. Before we can enter upon a detailed analysis of this matter, we must clarify the meaning of the term *taxonomic relationship* in the recent literature and especially our employment of this term in this book. The variety of meanings in which "taxonomic relationship" has been employed in the literature has led to confusion and misunderstanding among taxonomists. The meanings attributed to it seem to fall into two major classes. One use of the term is in a general sense in which the relationship may be phenetic, phyletic, cladistic, genetic, and, for that matter, on any other conceivable basis. The second meaning is more restrictive, including among "relationships" or "taxonomic relationships" only those usually called phylogenetic.

The first, and wider, defined use of the term is more prevalent in the literature and is the one adopted here. Among others, Simpson (1961, pp. 58, 62, 129, 130) and Michener (1957, p. 160; 1963, p. 154) both use taxonomic relationships broadly for phyletic as well as phenetic relationships. Acceptance of the broad definition for taxonomic relationships necessitates the use of a qualifying adjective to make the intended meaning more precise. Thus we speak of phenetic relationships, cladistic relationships, and so forth.

The confusion over meaning stems in part from the fact that in English the single term relationship symbolizes at least two separate concepts. *Webster's New Collegiate Dictionary* (1959) lists two definitions. The first is "the state of being related"; that is, "connected by reason of an established or discoverable relation." Thus there may be spatial, morphological, ecological, temporal, as well as genealogical relationships. The second is "Kinship; consanguinity, or affinity." In other languages these two concepts are identified by separate terms, hence their meanings cannot be confused. Thus, in German, the first meaning of relationship is *Verhältnis* or *Beziehung*, while the second is *Verwandtschaft*.

Our decision to retain the wider meaning of the word, accompanied by qualifying adjectives, is based on our reluctance to remove the term from the general scientific vocabulary. There appears no suitable synonym in the English language for the term in its first meaning and we can ill do without the general scientific and philosophic concept of "relationship."

In the past we have used "affinity" essentially as synonymous to relationship, qualifying it by adjectives to clarify the meaning intended (Sokal and Sneath, 1963). Cain and Harrison (1960b) have referred to phenetic affinity and phyletic affinity as synonymous with phenetic and phyletic relationship. It would appear that affinity carries the connotation of kinship, hence relationship by descent, in the minds of a great many biologists, for example, Haas (1962). We have therefore avoided using this term in the present volume.

Phenetic relationship has been defined as "... arrangement by overall similarity, based on all available characters without any weighting..." (Cain and Harrison, 1960b, p. 3). It is important to add that this measure of similarity does not carry with it any necessary implication as to relationship by ancestry, but does imply exhaustive estimates of similarity among phenotypes. Ghiselin (1966, 1969) has criticized the concept of overall similarity, believing it to be philosophically unsound. In a formal sense he may be correct: in fact such measures as actually used in phenetic taxonomy are measures "over some" rather than "over all," meaning measures over some sets of characters rather than over all available characters. It seems to us that the term phenetic similarity should be extended to mean similarity over any set and number of phenotypic characters. Whether the number of characters in an organism is finite and whether a measure of overall similarity, in the strict sense, is possible is discussed in Section 3.8.

Phenetic similarity based on restricted sets of characters may be of value in special evolutionary studies. Phenetic similarity of organs may reveal functional

or adaptational similarity (as shown in a study of wings of terns, gulls, and skimmers by Schnell, 1969). The resemblance of mouth parts of nectar-gathering insects may be related to the phenetic similarity of floral structures; similarity among larval forms in a group of organisms may be contrasted with that among adults and related to differences and similarities among their habitats (as illustrated in bees of the genus *Allodape*, C. D. Michener, personal communication).

Burt (1964) has raised the issue whether the term phenetic similarity should be restricted to similarity computed from equally weighted characters only (as in the original definition of Cain and Harrison). Regardless of the merits of equal weighting of characters (which we continue to maintain and shall discuss in Section 3.9) it would seem that phenetic similarity can be based on equally or unequally weighted characters as long as the operation for obtaining the similarity has been defined explicitly by the investigator.

Stemming from these considerations we may redefine *phenetic relationship as similarity (resemblance) based on a set of phenotypic characteristics of the objects or organisms under study*. While phenetic similarity may be an indicator of cladistic relationship it is not necessarily congruent with the latter. The magnitude of phenetic relationships between pairs of any set of objects depends on the kinds of characters and similarity coefficients employed.

The term phenetic relationship has given rise to other similar terms and combinations. *Phenetics* is that aspect of taxonomic relationship concerned with phenetic relations but it is also used as synonymous with the study of phenetic relationships. Combinations such as *phenograms* and *cophenetic value* will be encountered later.

Cladistic relationship was defined by Cain and Harrison (1960b) as relationship expressing recency of common ancestry. Cain and Harrison also speak of cladistic affinity to connote the degree of recency of ancestry. The term *cladistics* is used here to mean a study of the pathways of evolution; that is, how many branches are there, which branch came off from which other branch, and in what sequence? Cladistic relationship can be defined as—and represented by—a *branching (and occasionally anastomosing) network of ancestor-descendant relationships*. These treelike networks expressing cladistic relationships are called *cladograms* (Mayr, 1965; Camin and Sokal, 1965).

Any attempt to measure degree of cladistic relationships runs into several difficulties. One approach is that of Hennig (1966, p. 74) that “A species *x* is more closely related to another species *y* than it is to a third species *z* if, and only if, it has at least one stem species in common with species *y* that is not also a stem species of *z*.” This concept is illustrated in Figure 2-2. Hennig points out that by his definition species **B** would be phylogenetically closer to species **C** than it would to species **A**. Such a definition is strictly cladistic. Hennig adds that some phylogenetic systematists would object to such a definition because **B** is at a lesser distance from **A** than it is from **C** or **D** (in current terminology this is a phenetic relationship), while others would point out that the time (T_2) of the origin of **B** is closer to that of **A** (T_1) than to that of **C** and **D** (T_3). The latter is a chronistic concept of phylogenetic

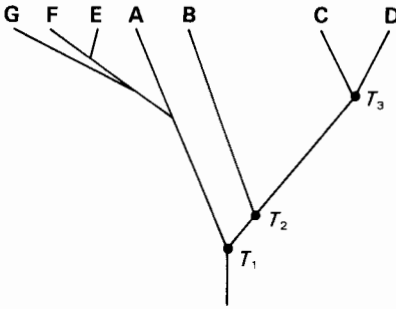


FIGURE 2-2

Cladogram (dendrogram of cladistic relationships) to illustrate the idea of cladistic affinity. Boldface capital letters represent species, T_i represents time. [Modified from Hennig (1957, Figure 5) and Sokal and Camin (1965).]

relationship. If the cladistic sequence of a group of organisms is known, Hennig's definition is operational. However, by the nature of the definition many cladistic relationships are undefined. Thus **B** is equally closely related (*sensu* Hennig) to **C** or to **D**. Since all species in this cladogram share a common ancestor at the first branch at time T_1 , species **B**, **C**, and **D** are equally related to **A**, **E**, **F**, and **G**.

Another plan would be to count the number of furcations that one would have to pass in order to go from one OTU to another in a cladogram. This is the cladistic difference discussed by Farris (1969b). By application of this criterion **D** in Figure 2-2 would be closer to **C** than it would be to **B**; **D** would be closer to **A** than it would be to **G**. However, this concept has some serious problems in practice, since an unknown number of branches would come off each stem, one for each extinct line.

Chronistic relationship refers to relationship between pairs of OTU's on the time scale of evolution. Frequently this is given as the vertical axis in a cladogram. Colbert (1963) has presented an extensive discussion of the time dimension in phylogeny. The relationship between phenetics and the time dimension is discussed in Section 6.1. Chronistic relationships of themselves are of limited usefulness. When combined with phenetic relationships they are of great interest for finding evolutionary trends and rates.

Phyletic or *phylogenetic relationship* are terms that have been employed in the past to include, usually in some undefined combination, the three types of relationships discussed above. Some authors, such as Davis and Heywood (1963), discuss phenetic and cladistic aspects of phyletic relationship separately. Others (Mayr, 1965; Simpson, 1961) consider all three elements (phenetics, cladistics, and chronistics) for establishing phylogenetic classifications. Yet others (following Hennig, 1966) would prefer phylogenetic relationships to be entirely cladistic. Thus the reader must be on his guard for different interpretations of the terms phyletic or phylogenetic.

Since phyletic relationship can have different meanings, it is desirable to replace it by the more precise terms defined above when precision is of the essence. However, there will frequently be instances when a more general, inclusive term will be useful, and we propose to employ the terms phyletic or phylogenetic relationship in this more general way.

2.4 PROBLEMS OF ESTIMATING PHENETIC RELATIONSHIPS

Numerous problems beset the taxonomist wishing to estimate relationships between taxonomic units. This is true whether phenetic or cladistic criteria are used. There are four major *problems of phenetic classifications* requiring discussion. They are (1) incongruence between classifications based on different parts of the body or different life history stages, (2) differences in estimates of relationships produced by different similarity coefficients, (3) differences in interpreting relationships produced by different clustering methods, and (4) the possible effects of parallelism and convergence on taxonomic judgments based on estimates of phenetic relationships.

Incongruence and Methodology

Incongruence between Classifications Based on Different Organs or Different Life History Stages. Incongruence between such classifications is related to the general problem of choice of characters discussed in Section 3.6. Phenetic similarity between pairs of taxonomic units will depend on the sets of characters chosen. It has therefore been realized that general classifications should be based on as broad a phenetic spectrum as possible, and that those classifications based on more restricted sets of characters such as from single organs and life history stages may have heuristic value for special problems.

Different Estimates of Relative Phenetic Similarity Depending on the Similarity Coefficient Employed for a Given Study. The general subject of such estimates, and similarity coefficients, will be taken up in detail in several sections of Chapter 4. It is not at all clear at this point that a unique measure of similarity between pairs of OTU's is possible or even desirable.

Different Classifications Resulting from Different Clustering Methods. Unless the sole aim of taxonomy is to reconstruct cladistic relationships (an approach we cannot recommend; see detailed discussion in Sections 2.5 and 2.6), there will be a variety of criteria for optimal classifications and it will be possible to reclassify the same similarity matrix in a number of ways so as to bring out different desirable classificatory aspects. Details of the problems of comparing different clustering methods and approaches to evaluating these are discussed in Section 5.10.

For purposes of the present discussion we might summarize our attitude as follows: while differences in sets of characters, similarity coefficients, and clustering methods will lead to different results in phenetic taxonomy, there are realistic expectations of finding commonly accepted solutions for frequently occurring classes of data and situations. Furthermore, some types of differences in results may themselves be of great interest leading to new insights into the nature of the organisms or of the relationships being studied.

Convergence and Parallel Evolution

The issue is often raised that the results of phenetic taxonomy are likely to be erroneous (by the criteria of phylogenetic taxonomy) because of the possibility of serious discrepancies between the relationships based on similarity and those based on descent. It is almost a truism that an intimate relation must exist between phenetic evidence and the degree of relation by ancestry. It should also be obvious that, while the two kinds of taxonomy are equally valid for their own purposes, they stand in a peculiar relationship to each other: if knowledge about phyletic relations is required, it must be inferred from phenetic evidence; phenetic relations, however, are inferred not from phyletic hypotheses but from the specimens themselves. A ready analogy offers itself here: we may estimate the similarity between geometric objects on the basis of the nature and size of their dimensions, without any implication as to their past history or how they were developed or constructed by geometers. Yet under some conditions we may *infer* from their geometric form certain probabilities about their past history. We may, for example, suggest that a regular octahedron was developed by geometers subsequent to the development of the square.

Convergence and parallelism are terms over which there is considerable confusion (see Haas and Simpson, 1946, for a full discussion), since they may mean convergence or parallelism in one organ (or in one character complex) or of the entire phenotype. Many authors do not specify which they mean. In this book we use the terms with reference to phenetic similarity, whether the changes are contemporary (isochronous) or not (heterochronous). Convergence restricted to a small part of the phenotype would not, of course, produce convergence in overall similarity; for example, the bats are convergent on the birds with respect to flight, but in the remainder of their phenotype they are divergent from birds, when compared with the reptilian ancestor both have in common.

Convergence. This is no problem at all when considered strictly by the criteria of phenetics. So long as we are concerned with phenetic relationships the similarity value obtained by numerical methods will be representative. It is only when we wish to draw *phylogenetic* conclusions—rather than phenetic ones—that convergence may confuse the issue.

It is necessary to specify in what respects lines are convergent. It is quite possible for two lines to converge in respect to one organ and to diverge in respect to others. The only kind of convergence that is pertinent to our present argument is that where the lines converge in so many respects (that is, characters) that it causes an increase in the overall similarity of the two lines. This, which can be called “overall convergence” to distinguish it from convergence in a few respects, might cause serious discrepancies between the taxonomy yielded by phenetic methods and a classification based on a cladogram. Convergence in a few respects (“organ convergence”) will not cause discrepancies, since these few respects will have little effect on the

resemblance of the many nonconvergent character states included in the analysis. There is of course no sharp line between overall convergence and organ convergence.

The pertinent question, then, is: Does marked overall convergence ever occur? And it is one which urgently needs study. There are many examples in which numerical taxonomy could be readily employed to test this question: *Canis* and the thylacine marsupial wolf, the marsupial and eutherian moles, the seals and sirenians, and some xerophytic or parasitic plants, could all be compared. We believe the overall similarity of pairs of this kind is not high. Indeed, if this were not so, it is uncertain how they were recognized as "convergent pairs" and not just close relatives. The very obvious and striking similarities in appearance will, we believe, account for very few of the total features analyzed. Any reasonably random and unbiased selection would, we think, include far more features which did not show convergence. And even in those systems whose adaptations seem primarily responsible for the observed convergence one finds many differences due to different modes of achieving the same function with diverse anatomical parts. We know of no cases of striking overall convergence where the phylogeny is thoroughly known. Sneath (1961) has pointed out the ridiculous implications of total overall convergence in higher animals, which may not be so ridiculous in viruses (see Section 2.6). It is perhaps worth noting that when considering many characters there is every expectation that evolutionary processes will *overall* be divergent. This is a consequence of variation, which has a strong element of randomness. In order to obtain convergence, the possible kinds of variation must be restricted; that is, there must be more change in different features than in common features. We have no reason to believe that, in general, natural selection will have this effect (except possibly in situations like that described below), since it will act largely upon random mutations and in many different directions.

What we have said above applies to the higher levels of taxonomic rank. At lower levels, at the genus and below, there is a possibility of some degree of overall convergence. For example, the introduction of a new food plant into an archipelago possessing several island races of a fruit-eating bird (which had diverged slightly from one another over the course of time) might produce such a strong selection pressure—in the same direction and in all the islands simultaneously—that these races would rapidly evolve toward adaptation to feeding on the new plant, and this might outweigh the slow accumulation of genetic differences that had been continuously occurring in each race. The overall similarity between these races might then increase somewhat, resulting in some degree of overall convergence. In the absence of knowledge of the past, it is difficult to see how any systematic procedure would elucidate the case, and numerical taxonomy is in no worse position than others. It is possible that convergence on this level and of this degree may be frequent but undetectable, at least by any of the procedures known today.

Finally, we may emphasize an obvious but often forgotten point. If we do indeed wish to study convergence, we can only do this by comparing a cladogram with the

phenetic arrangement through time. In no other way can we detect the process of convergence, and any attempt to restrict taxonomy to cladograms would defeat its own ends.

Sokal and Camin (1965) have pointed out that most cases termed convergence in conventional taxonomy would not be so considered in numerical taxonomy. When characters are recognized to be only *superficially* similar they would have to be coded as different characters and therefore would not contribute to the measure of similarity, but will in fact reinforce divergence once the characters are coded differently. For example, a detailed comparison of the wings of birds and bats may show them to be less similar to each other than either is to be a generalized tetrapod forelimb. As soon as we go into sufficient structural or physiological detail, convergent characters or processes would resolve themselves into different ones because of their diverse origins. Examples in point would be similar structures produced in different phylogenetic lines (the eyes of vertebrates and cephalopods are similar, but important embryological, histological, and structural differences would cause their characters to be coded differently). Without any knowledge of the underlying physiological mechanisms we would code known instances of DDT-resistance in various insects uniformly, but we would possibly code them quite differently if the physiological mechanisms conferring resistance on the insects were known. If similar characters or their states are not recognized to be different in taxa of diverse phylogenetic origins (that is, the convergence misleads the taxonomist), or if in fact the characters are identically convergent, the characters would be considered operationally homologous (see Section 3.4) and the states scored identically. In the first of these cases the characters would be miscoded and would erroneously contribute to phenetic similarity. In blue insects whose color is due to different causes (pigment and optical interference), we may—in ignorance of the true nature of the colors—code the insects identically as “blue” by operational homology. The similarity due to this common coding is the error in evaluating their phenetic similarity and is removed as soon as the state “blue” is recoded appropriately.

Sokal and Camin (1965) emphasize that for certain *special* classifications it may be desirable to leave such miscoded characters identically coded, since it is necessary to study nature at different levels of organization and complexity. An understanding of life processes at the finest level does not necessarily contribute to an elucidation of processes at a higher level of organization. Different phenetic relationships may therefore be appropriate for different levels of organization. Thus, when discussing problems of population or ecological genetics, the resistance of insects to an insecticide may be the significant feature, although it may be erroneously coded in terms of the physiological mechanism of the resistance. In a study of their physiology, the mechanisms by which these insects are resistant become the feature to be coded and correct homologies may be established. In a pharmacognostic classification of plants the presence of a specific alkaloid may be

coded identically, although it may be known to arise through different metabolic pathways.

Parallel Evolution. This occurs when two genetically isolated stocks evolve so as to keep constant the difference in those attributes which are under consideration. Parallel evolution seems to us to be similar to convergence and subject to analogous reservations of definition, but to a lesser degree. The same problems and dilemmas arise, and again there is very little clear evidence in favor of extensive parallel evolution if all the features are included in the taxonomic analysis. Possibly the best instances of what may be *overall* parallel evolution (the inclusion of the word "overall" implies a constancy in the overall similarity, as with convergence discussed above) occur in certain ferns (Holttum, 1949) and also in certain ammonites. The apparently parallel trends in the degree of convolution and ornamentation of the shells of ammonites represent very few characters out of the many present during life. In addition to this, these characters are likely to be selected by the environment in the same way; for example, it is possible that certain forms of the shell may have protected many different species of ammonite from a particular predator. Another example of this may be the repeated evolution of increased curvature of the shell in lineages of oysters, leading to the *Gryphea* phenotype, which has been interpreted as a recurrent adaptation to a muddy sea bottom (discussed by Joysey, 1959). This may well have involved only a few characters, since we do not know what changes occurred in the soft parts of these molluscs.

The reviews of Trueman (1930), Swinnerton (1932), George (1933), and Joysey (1959) may be consulted for some of the better-known instances of this phenomenon; these authors all emphasize the difficulty in deciding whether the apparently parallel lineages are indeed independent phyletic lines, or whether the forms found in any given stratum should be grouped together in a monophyletic taxon. If the latter procedure is correct, then parallel evolution is simulated by similar adaptive radiations in successive taxa (stages) of an evolutionary line.

As with convergence, most apparent examples of parallel evolution are due to parallel trends in a few characters. Again, quantitative studies of this problem are urgently needed. Even if overall parallel evolution does occur, it will be no easy matter to prove convincingly the validity of the phyletic lineages concerned, and the phenomenon will probably be of small degree.

Readers of the above account may feel at variance with our ideas because our definitions of convergence and parallelism, while not alien to the usage of systematists, are probably not the most common ones; hence our comments may be thought not relevant to the central ideas of convergence and parallelism as customarily conceived. While the terms convergence and parallelism permeate evolutionary literature, it is difficult to find definitions for them. Remane (1956), for example, uses but does not define convergence and parallelism. Rensch (1947) considers convergence to be simply nonparallel evolution. Simpson (1961, p. 78) has

attempted to coin precise definitions for these terms. He defines parallelism as “the development of similar characters separately in two or more lineages of common ancestry and on the basis of, or channelled by, characteristics of that ancestry.” Convergence he defines as “the development of similar characters separately in two or more lineages without a common ancestry pertinent to the similarity, but involving adaptations to similar ecological status.” The inferences that are customarily made from such a definition, and that are also discussed by Simpson (1961, pp. 103 ff.), are that convergence occurs between forms that are relatively far apart, while parallelism occurs only among lines that are relatively closely related. These also are the meanings ascribed to these terms by Mayr (1969a). These definitions are useful in a general way for describing evolutionary phenomena. It is useful to have a term to describe the superficial similarity of fishes and whales, for example, and to distinguish it from the independent acquisition of similar coloration in several closely related species of insects living in deserts. However, these definitions would have little operational value even if the complete evolutionary history and genetic structure of the taxa in question were known. They are of even less value in the more realistic case where such data are not available. Simpson mentions the occasional literal interpretation of convergence as the narrowing of differences between lines and the parallel change of these lines as constituting parallelism, but he does not consider such an interpretation to be a particularly useful taxonomic concept.

It seems to us, on the contrary, that in any taxonomy based on a phenetic system (in fact, any taxonomy one wishes to base on measurable quantities), the definition of parallelism and convergence should be entirely based on the parallel or convergent nature of the differences between the lines. We hold these beliefs for several reasons.

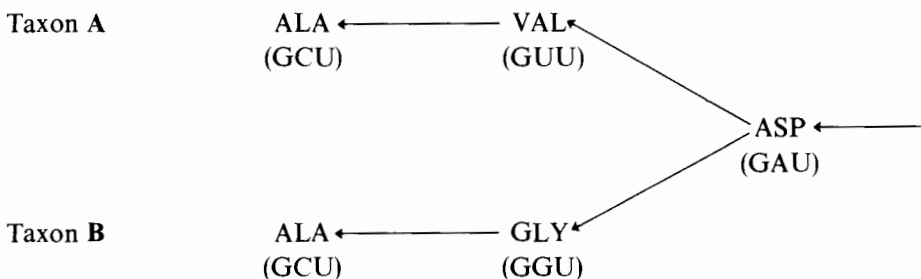
1. No fundamental and useful distinction can be established between convergence and parallelism *sensu* Simpson and Mayr. To say that convergence takes place only between distantly related forms prejudices the issue of relationship completely. We would no longer be able to use cases of convergence or parallelism in our classificatory schemes because relationships would have been predetermined before any decision could be taken on whether a convergence or parallelism is at hand.

2. The definitions of Simpson (1961) and Mayr (1969a) and their subsequent discussions lead one to believe that by convergence is meant the construction of similar structures based on different genetic systems, and that by parallelism is meant the construction of similar structures based on similar genetic systems. However, these assertions are based on inferences from the systematic positions of the organisms involved and lead to the same difficulties noted under 1.

3. In the absence of knowledge about genetic homologies, we are faced in the main with phenetic changes that may be divergent, parallel, or convergent. The only useful distinction is whether changes are parallel or convergent, or better still

whether there is relatively more or less divergence, since often less divergence between some members of two taxa than between the majority of the members of the taxa may be considered as convergence or at least parallelism. Of course, *sensu strictu*, this can only be done if fossil series of organisms are available. Where only living end points of evolutionary change are known, parallelism or convergence cannot be shown but must be inferred with much uncertainty from recent evidence.

4. It must also be stressed that at least at the molecular level true phenetic convergence can indeed take place. If the following is the true cladogeny of a protein position in two taxa



then the two alanines are identical regardless of their different origin from aspartic acid via valine and glycine, respectively. With respect to this protein position the two taxa are in fact identical; and the DNA representation of these two alanines might also be identical, as shown above by the nucleotide triplets in parentheses.

Components of Phenetic Similarity

We shall conclude our discussion of phenetic relationship by reviewing two components that combine to yield phenetic similarity: similarity due to common ancestry is called *patristic similarity* (Cain and Harrison, 1960b) and that due to parallelism and convergence (true or miscoded) is *homoplastic similarity* (Simpson, 1961). Patristic similarity is in general less than total phenetic similarity, since it consists of the latter minus the homoplastic component, as pointed out by Cain and Harrison.

It is evident that cladistic relationship is not directly related to the above two types of relationship. They are phenetic, and measure similarities between the forms. Cladistics is concerned with the way in which the lineages branched and not with the degree of difference. A genealogy or pedigree is an example of a cladistic scheme.

Patristic similarity can be further subdivided; the present analysis differs in some respects from an earlier one by Sokal and Camin (1965). One component of patristic similarity may represent *primitive similarity*, due to character states that are identical in two descendent taxa and identical with the states of the characters in their most immediate common ancestor. In Figure 2-3 both lineages C and E have diverged from a common ancestor A. The states of the two characters 1 and

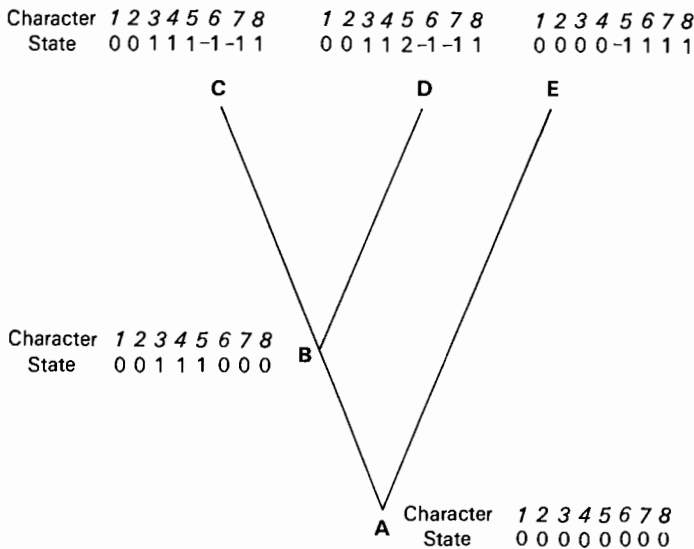


FIGURE 2-3

Cladogram showing three extant organisms, C, D, and E, and two ancestors, A and B, with eight characters, to illustrate patristic and homoplastic components of phenetic similarity (see text).

2 have not changed in the two descendants from the states that occurred in the ancestor. This represents an element of primitive similarity with respect to homologous character states between the descendants and also between either descendant and the ancestor. Hennig (1966) calls this component of patristic similarity *symplesiomorphy*. Our term is simpler and of greater mnemonic value.

The other component of patristic similarity is *derived similarity*. This results from character states that are the same in the descendants but not identical with those in a more remote common ancestor, as illustrated by characters 3 and 4 in lineages C and D in Figure 2-3 with respect to the ancestor A. This concept is only applicable when the common ancestor is more remote than the most immediate one, because C and D show primitive similarity for characters 3 and 4 with respect to their immediate ancestor B. These characters became modified to state 1 in intermediate ancestor B. The contribution of the two characters 3 and 4 to the similarity between C and D is the same as that for characters 1 and 2 in lineages C and E, but 3 and 4 contribute less to the similarity of C or D with A than to that of E with A. Hennig (1966) calls this relationship *synapomorphy*. Derived similarity includes cases where the states in the two descendants are not identical because of further evolution in one of the lineages. An example is character 5 in lineages C and D. In the latter the character evolved to state 2 from the ancestral state 1 in B. But C and D are relatively more similar to each other than either is to E, which has evolved state -1 for character 5. Derived similarity should be differentiated conceptually from primitive similarity, the other component of patristic similarity. If

the phenetics of the ancestors are known, the primitive similarity can be separated from the derived similarity. Character states that exhibit derived similarity are generally considered to be homologous in the classical sense of the term.

Another evolutionary phenomenon is illustrated by characters 6 and 7, which evolved independently in lineages C and D, from the ancestral state 0 in A and B to the derived state -1. The evolutionary phenomenon involved is clearly parallelism as generally understood. Another term for similarity due to parallelism is *homoiology* (L. Plate, cited by Hennig, 1966). Character 8 evolved three times to the derived state 1 in lineages C, D, and E, and in this case it illustrates convergence of C and D on E with respect to A and B.

With respect to characters 3 and 4, taxa C and D in Figure 2-3 are both divergent from A, although C and D are identical for these characters (showing derived similarity with respect to A, and primitive similarity with respect to B). With respect to character 5, taxa C and D are again divergent from A (D is also divergent from B), but if we assume that the character must evolve from state 0 through state 1 before reaching state 2, then taxa C and D show derived similarity as well as divergence with respect to each other and A.

Statements about amounts of phenetic similarity in Figure 2-3 are only relative. Thus we suppose that the phenetic similarity between taxa C and D will be greater than between C and E. The amount by which this will be true depends on the measure of similarity employed.

Finally, we should point out that Figure 2-3 does not necessarily imply that the evolutionary step from state 0 to state -1 in characters 6 and 7 and to state 1 in character 8 occurred simultaneously in both lineages. If it occurred early in the line B to C and late in the line B to D (heterochronic parallelism, Maslin, 1952), we could at the appropriate points in time speak of a divergence of the line BC from BD, followed by a convergence of C upon D. Divergence and convergence are here employed as phenetic concepts. This meaning of convergence differs from the one used conventionally, although some taxonomists employ the term phenetically without clearly defining it. Heterochronic parallelism would seem to be a preferable term. Since primitive and derived similarity result from common ancestry their sum yields the overall patristic similarity.

The homoplastic component of phenetic similarity represents those components that show parallelism and those that are convergent. Homoplastic similarity is thus due to homoiology as well as to truly convergent characters; that is, those that are identical but of different ancestry, like the protein position discussed above. It also expresses similarity due to characters or their states, or both, that have been misinterpreted or miscoded in a convergent manner. Recognition of homoplasia is important for two reasons. Once we recognize that a character has been miscoded we are led to revise and correct our coding system. Also, homoplasia suggests that similar evolutionary forces are at work in the lines under study, which tend to make some of their characters resemble each other. This is a phenomenon of general

biological interest. The analysis of phenetic similarity between two taxa with respect to a given ancestor more remote than the most immediate one can be summarized as follows:

Phenetic similarity = Patristic similarity + Homoplastic similarity

Patristic

similarity = Primitive similarity + Derived similarity (together these represent classical homology)

Homoplastic

similarity = Similarity due to parallelisms (homoiology)
+ Similarity due to convergence (identical characters or states, or both, derived from different phyletic lines, or different characters or character states erroneously coded as homologous).

Hennig (1966, p. 146) has independently arrived at a similar partitioning of similarity into symplesiomorphy (= primitive similarity), synapomorphy (= derived similarity), and convergence (= homoplastic similarity).

We should point out that when taxonomic relations are expressed as dissimilarities, phenetic dissimilarity equals patristic dissimilarity decreased by the similarity due to homoplasy.

There are other ways of partitioning phenetic similarity, discussed in Section 4.11. One of these is the distinction between size and shape.

In conclusion it may be noted that many of the difficulties discussed in this section are due to our uncertainty of what kind of similarity is most appropriate in taxonomy, and this is more of a problem of the taxonomist than of the statistician.

2.5 PROBLEMS OF ESTIMATING CLADISTIC RELATIONSHIPS

With the advent of the Darwinian theory of evolution the natural classifications of earlier taxonomists became explicable in terms of descent from a common ancestor: a taxon was now interpreted as a monophyletic array of related forms. It has, however, been frequently pointed out (as by Bather, 1927, and Remane, 1956) that this change of philosophy did not bring with it a change in method. Taxonomy proceeded as before; only its terminology had changed. Remane (1956) quotes Naef (1919, pp. 35–36):

... und was Haeckel und die Phylogenetiker zunächst getan haben, war nichts anderes als die Übersetzungen der speziellen Einsichten, die sich an diese Lehre früher geknüpft hatten, in eine Sprache durch Anwendung einer neuen Terminologie, ohne doch die Lehre selbst einer Vertiefung zuzuführen oder einer kritischen Betrachtung zu unterwerfen. Auch die—wenig abgeklärten—Grundbegriffe der alten Morphologie wurden

von Haeckel einfach in die neue Sprache übersetzt, die dem Wesen nach eine genealogische war. Dabei wurde dann

aus Systematik	Phylogenetik,
aus Formverwandtschaft	Blutsverwandtschaft,
aus Metamorphose	Stammesentwicklung,
aus systematischen Stufenreihen	Ahnenreihen,
aus Typus	Stammform,
aus typischen Zuständen	ursprüngliche,
aus atypischen	abgeänderte,
aus niederen Tieren	primitive,
aus atypischer Ähnlichkeit	Konvergenz,
aus Ableitung	Abstammung usw. usw.

Freely translated, this reads :

... and what Haeckel and the phylogeneticists did at first was nothing more than to translate the special insights which had been gained in this field of study into another language by applying a new terminology without any deeper analysis or a more critical examination of the concepts. The poorly defined fundamental concepts of the old morphology were also translated by Haeckel into this new, essentially genealogical language. Thus

Systematics	became phylogenetics
Relationship by form	relationship by blood
Metamorphosis	evolution of phyletic lines
Systematic series	ancestral series
Types	archetypes
Typical states	primitive states
Atypical states	derived states
Lower animals	primitive ones
Atypical similarity	convergence
Derivation	descent, and so on and so forth.

In view of the conclusions of the previous sections it is obvious that phylogenetic relationships need to be subdivided and refined into phenetic and cladistic relationships. Classifications based on "phylogenetic relationships" without this clear separation and definition will be ambiguous and of little value to either phenetic or cladistic taxonomists. Hennig (1966, p. 77) states this position well: "... combining different systems in a syncretistic system robs the combinations of any scientific value, since it could never be used as the basis for investigations that presuppose knowledge of a particular kind of relationship between the organisms." The necessity for separating phenetic and cladistic aspects of taxonomic relationships and of basing classifications on only one or the other aspect are recognized by most modern taxonomists who have given thought to this matter.

If cladistic relationships are to be the basis for erecting classifications, we must inquire into the possibility of estimating such relationships with a reasonable

degree of reliability. Supposedly cladistic arrangements have been published innumerable times in the systematic literature ever since the publication of *The Origin of Species*. There are few biological topics on which more misinformation has been published. Even in paleontology the ratio of fact to speculation is rather low. "Works which refer to the fossil evidence of evolution usually cite a few of the well-known cases of evolutionary series as if they were merely representatives of a host that might have been quoted, instead of stressing the fact that records of such cases are rare" (Challinor, 1959).

The critical question to be answered is: Can cladistic lineages be estimated from information available to systematists? Hennig (1966), for whom the determination of cladistic sequences is the crucial problem in taxonomy because of his advocacy of strictly cladistic methods of classification, presents by far the most detailed and carefully reasoned discussion of this issue in the modern literature. The criteria considered by Hennig include consideration of problems of homology, character phylogeny, reversibility, convergence, and parallelism, as well as a discussion of chorological and chronological character variation. While Hennig prefers to maintain a phylogenetic definition of homology, he realizes correctly that the actual recognition of homologies has to be through character correspondences and his ideas for determining homologies in practice are thus very close to the concept of homology espoused by us and discussed in Section 3.4. However, as is shown there, this concept is generally phenetic and cannot be employed for rigorous cladistic analysis.

Hennig's section on character phylogeny (1966, pp. 95 ff.) is the heart of his discussion of the use of morphological evidence for cladistic analysis. These considerations are fundamental in cladistic work, whether it is done by the more traditional methods, by Hennig's system, or by the operational techniques of numerical cladistics discussed in Sections 6.3 and 6.4. To satisfy his criterion of monophyly, Hennig feels that only synapomorphous relationships (derived similarity) between species can assist in the establishment of monophyletic lineages. This means that one must distinguish primitive from derived character states. Among the criteria cited by Hennig that would aid in this undertaking is geological evidence. Character states found in earlier geological strata can be considered primitive when compared to character states found in later geological strata. Although such considerations may be useful in an approximate general way, they do not sustain critical scrutiny. The only reason we compare forms in these two strata is because they are similar to each other. Thus we use the phenetic similarity of putatively related forms to make judgments on whether a character in question should even be compared. Consequently, a decision from geological evidence, on whether a character state is primitive or derived is fundamentally based on the phenetic similarity of the forms studied. It might be argued that with extensive fossil series the ordering of the character states would be conclusively evident, but again it should be pointed out that the only way to establish these series



is by the phenetic similarity of the forms under consideration. We are left with what is essentially a character analysis in a phenetic numerical taxonomy of these organisms (see also Colless, 1967b, and Section 3.4 for a discussion of this point).

An identical argument holds for chorological progression series. Such series are formed by species that displace each other in a geographical progression and exhibit primitive characters in the area of origin and derived characters in the most recently invaded territory. Yet again the decision on whether one should even compare these related species, or the very fact that these species are related, is based on phenetic considerations.

A third criterion for ordering character states is based on Haeckel's well-known law of recapitulation by which the successive embryonic stages of an animal are said to mirror its phylogeny. It is now realized that the many exceptions from this law can lead to serious errors of interpretation. Hennig (1966, p. 96) concludes that "the transformation stages that led to a character condition during phylogeny, ... cannot be read off with certainty from the ontogeny of the species," although "the 'criterion of ontogenetic character precedence' remains an important aid in phylogenetic systematics, provided it is not uncritically evaluated more highly than other aids."

Hennig's fourth criterion is his strongest. He claims that under certain circumstances knowledge of the evolutionary trends of the states of one character may yield information on the evolutionary order in another character. Figure 2-4 illustrates this concept. Character h has evolved from primitive state h_0 , which is retained in species **A**, to state h_1 in species **B** and state h_2 in species **C**. Paraphrased into the terminology of this book, Hennig states that if the evolutionary order of the states of character h is known, the conclusion that the species **B** and **C** are most closely related is compatible only with the assumption that in the second character i with three character states i_0 , i_1 , and i_2 , the primitive condition is represented by i_0

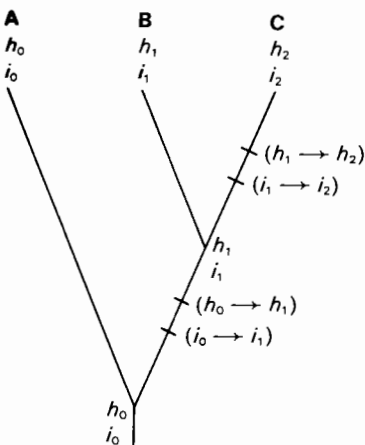


FIGURE 2-4

The three species **A**, **B**, and **C** have evolved from a common ancestor which possessed the character states h_0 and i_0 for characters h and i respectively. At points in evolution shown by the cross bars the states evolved into other states (as shown by the transitions indicated by arrows and given in parentheses). [Modified from Hennig (1966).]

and the most strongly derived condition by i_2 . However, this statement does not meet the real issue. In any given case we are unlikely to know the true cladistics of the species in the manner in which it is shown in Figure 2-4. All that we would have is an hypothesis about the order of the states of character h and from this we can reconstruct the most parsimonious cladogram (Section 6.4) assuming that our coding of character h is correct. If in fact we were to know the cladistics of the three species, then the whole argument is beside the point. We would be little interested in evaluating the character states of i since in *Hennig's taxonomy* (as distinct from evolutionary studies) the main purpose of knowing which are primitive and which are derived is in being able to reconstruct the cladistic relationships of the three species.

To make the situation more realistic we should assume that we do not know the cladistics of the three species. Then what one needs to do is to try to fit various interpretations of the order of the character states i_0 , i_1 , and i_2 to possible cladistic arrangements compatible with our interpretation of the character states of h . In Figure 2-5 we can see that there are three possible cladistic arrangements of the three species (assuming that they have to be connected and do not issue in a single trifurcation from a common ancestor). To decide which of these cladograms is more likely we must make certain ancillary assumptions. The most obvious assumption is parsimony of evolutionary steps (changes in character states). Since in a realistic situation one is hardly ever certain about the evolutionary order of any one character, such as character h in this example, one must jointly determine the parsimony of the two characters together. On examining Figure 2-5 we find that, depending on how we code character i , either cladogram b (the one indicated as correct in Hennig's original figure) is the most parsimonious, or e and f are equally parsimonious. And these two are not necessarily the same, when different assumptions are made about the order in which the character states are permitted to change. Considerations of this sort led Camin and Sokal (1965) to their method of cladistic analysis, which jointly considers numerous characters coded according to assumed evolutionary sequences. The details of this method and philosophy will be taken up in Sections 6.3 and 6.4.

The point to be made here is that these methods, when done with many characters and when the possible alternatives are evaluated by computer, may lead to reasonable reconstructions of the cladistic sequences of the species under study; yet they cannot invariably lead to precisely constructed cladograms for the following reasons: (1) With more than three species and more numerous character states the number of possible solutions to be tested becomes very large. It is not obvious from Hennig's writings whether he is aware of the very large number of possible cladistic solutions in any phylogenetic scheme. It would be totally impractical to test all of these even on a computer, let alone by traditional procedures. (2) The methods chosen are more in the nature of overall compatibility tests of characters. It is conceivable at least in theory that uniformly miscoded characters could be

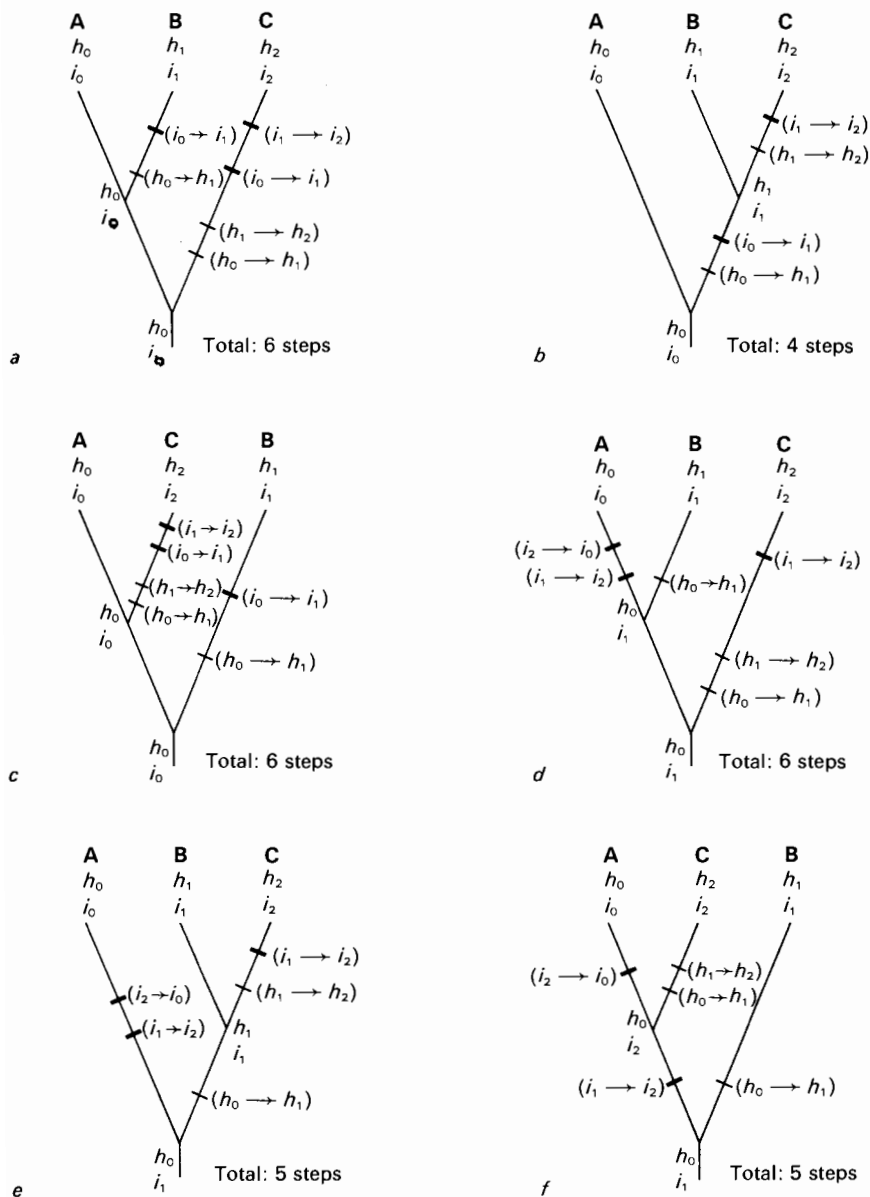


FIGURE 2-5

The three possible cladistic arrangements of the species A, B and C from Figure 2-4 are shown in a, b, and c, and in d, e, and f (omitting the case of divergence of all three from a single point). In a, b, and c the assumptions about the evolution of the characters h and i are that they followed the order of the subscripts 0, 1, and 2 in both cases. The most parsimonious solution is then cladogram b with four evolutionary steps.

In d, e, and f it is assumed that the order i_0, i_1, i_2 , postulated above, is incorrect, and using new assumptions that character i evolved $i_1 \rightarrow i_2 \rightarrow i_0$ it is seen that cladograms e and f are equally parsimonious with 5 evolutionary steps each.

mutually totally compatible and yet be entirely incorrectly interpreted. (3) Hennig correctly asserts that certain interpretations of a character *i* would be incompatible with other previous interpretations of character *h*. These considerations, which have been independently developed by Wilson (1965) and termed a consistency test by him, serve mainly to exclude impossible or improbable interpretations but cannot serve to confirm assumed interpretations of character *h* or character *i*. These views are stated more fully and firmly by Colless (1966, 1967b).

Among other criteria for evaluating the phylogenetic order of character states, Hennig mentions the supposed correlation between the character states of hosts and parasites. The pitfalls in this field are obvious and even greater than those encountered by matching the states of one character with cladograms produced from another character.

We can conclude with Hennig (1966, p. 146) that "there are criteria and rules for determining whether the characters of different species belong to a transformation series, and for deciding whether they are to be evaluated as plesiomorphous [that is, primitive patristic] or apomorphous [or derived patristic], *but these do not have absolute validity*" (our italics and bracketed synonyms). Colless (1967b) shows that without resort to phenetic principles Hennig's attempts to recognize plesiomorphous from apomorphous character states are unsuccessful. He concludes "... that the 'Hennig System' is simply an intuitive, prototypical form of statistico-phenetic taxonomy, and that its advocates simply misunderstand the nature of the latter process." This claim has been disputed by Schlee (1969) and defended in a rebuttal by Colless (1969a,b).

Other authors are even less explicit in furnishing reliable criteria for the recognition of the evolutionary interpretation of sequences of character states. A single quotation from Simpson (1961, p. 102) may suffice: "Perhaps the most conclusive evidence as to primitive (and hence ancestral) characters is provided when one condition in a group or one end of a sequence has a homologue in another group of more remote common ancestry." It is obvious that without knowledge of the evolutionary history (presumably as evidenced by the fossil record) we could make little progress here. Remane (1956) suggests recognizing cladistic relationships through the recognition of homologues, for which the main criteria are morphological correspondences. Remane's methods therefore cannot separate phenetic and cladistic relationships.

If it is difficult or impossible to estimate cladistic relationships from primary evidence (characters of the taxonomic units), we might ask whether cladistic deductions can be made from overall phenetic similarities between organisms. Let us take a closer look at the reasons for wishing to develop taxonomies that are in accord with phylogeny. This ideal is expressed mainly by the dictum that taxa should be monophyletic groups. There will therefore be difficulty only when the taxa given by phenetic taxonomy are not monophyletic but polyphyletic. We believe that numerical phenetics will in general give monophyletic taxa because

we believe that phenetic groups are usually monophyletic. To contradict this belief we must have evidence that phenetic groups, created by adequate and acceptable numerical techniques, are not monophyletic. A clear example would be where convergence has occurred to an extent that causes confusion. But we must first show that this degree of convergence has indeed occurred. Figure 2-6 shows that there are several interpretations of what is at first sight a simple problem. There may indeed have been convergence so that organisms **A** and **B** are more similar phenetically than their ancestors **A'** and **B'**, and **B** is convergent on **A**, though by

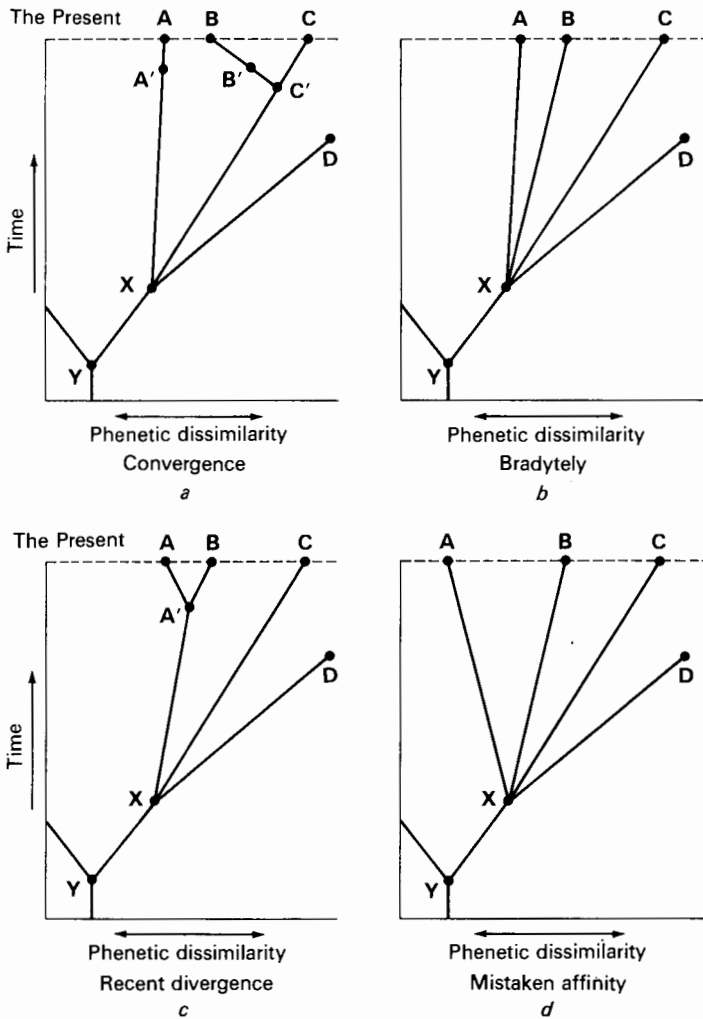


FIGURE 2-6

Alternative interpretations of apparent convergence. For explanation, see text.

ancestry related more closely to organism **C**; see Figure 2-6,*a*. But this conclusion may be uncertain. The evolution may have occurred as in Figure 2-6,*b*, where there is no convergence but divergence at different rates, with the phyletic line **B** evolving slowly (bradytely). Or, **B** may in fact have descended from **A'** so that **A**, **A'**, and **B** are monophyletic; see Figure 2-6,*c*. Finally we may have been mistaken in thinking that **B** was convergent on **A**, for a careful estimate of resemblance may show, taking all their attributes into consideration, that **B** is more similar to **C** than to **A**; here we had been misled by some striking but superficial or restricted set of characters. As soon as the homoplastic similarity due to these characters is removed the greater phenetic relationships of **B** to **C** becomes evident, as in Figure 2-6,*d*. It is clear that from a consideration of the organisms **A**, **B**, and **C** alone (without the evidence of the fossil forms **A'**, **B'**, **C'**, and **X**), we cannot distinguish between these alternatives except to recognize the last of them. Indeed, we have as yet no acceptable evidence that convergence of this kind—that is, overall convergence in phenetic resemblance—does take place to any marked extent. If it did, it would be exceedingly difficult to prove, for we would have to have an excellent series of fossils to be certain that we had not made any mistake in reconstructing the cladogram. The known examples of convergence are all open to the objection that relatively very few characters are affected.

Discrepancies between phenetic and phyletic taxonomies can also occur without convergence—for example by divergence at different rates, as in Figure 2-6,*b*, where **A** and **B** are more similar phenetically than either is to **C**, although cladistically all three are equally related.

Even if we grant that overall convergence can occur (as, for example, has been suggested for some groups of birds), we must ask ourselves why we should wish to make taxonomies based on monophyletic groups. Suppose the convergence had become so absurdly extreme that the two forms are almost indistinguishable and can readily and successfully hybridize: what is the purpose of separating them on grounds of ancestry when in all other attributes they are virtually the same? The purpose cannot be to emphasize minor dissimilarities, nor to serve as a guide to the behavior of these forms with respect to their genetic properties, or any other class of properties. The purpose is presumably, therefore, to show that this convergence had occurred, a fact which could be expressed in simple terms without any need for setting up the whole apparatus of formal systematics, and, as is discussed above, the classification would thus be for this special purpose, and not a general classification. Jardine (1969b) points out that in a natural classification monophyly can be regarded only as a diagnostic criterion, not as an absolute criterion for establishing the system. Where independent phyletic lines fuse into one, the whole problem becomes thoroughly confused, whether we know to expect confusion or not, since there are so many alternative ways of dividing the network of phyletic lines. This is a common occurrence in plants, through the mechanism of allopolyploidy.

Simpson (1961, p. 120) has pointed out correctly that most definitions of monophyly are nonoperational because "they are so vague that they provide insufficient criteria for separating one from the other [monophyly from polyphyly] by analysis of evidence." We can ignore the naive statements of earlier authors defining monophyly as descent from a single pair of progenitors, statements made in ignorance of contributions of modern evolutionary theory. Hennig (1966, p. 73) defines a monophyletic group as "... a group of species descended from a single ('stem') species, and which includes all species descended from this stem species. Briefly, a monophyletic group comprises all descendants of a group of individuals that at their time belonged to a (potential) reproductive community, i.e., to a single species." He points out the sometimes neglected fact that the monophyletic group has to include *all* species derived from this ancestral species, not only those which the taxonomist wishes to classify. As Simpson correctly states, the problem with a definition of this sort is that it is difficult to know how far back one has to trace separate stems in order to arrive at the common stem form. For example, by Hennig's definition it might well be that the mammals could be made monophyletic only by tracing them back to an unknown early reptile stem.

A useful analysis of Hennig's concept of monophyly has been undertaken by Ashlock (1971). He distinguishes between two kinds of monophyletic groups—those that are *holophyletic*, wherein all the descendants of the most recent common ancestor are contained in the taxon (this is a monophyletic group *sensu* Hennig), and those that are *paraphyletic* and do not contain all of the descendants of the most recent common ancestor of that group. A *polyphyletic* group according to Ashlock is one whose most recent common ancestor is not cladistically a member of that group. Because most if not all taxa presumably have common ancestors when one is willing to trace back their ancestry far enough, polyphyly in Hennig's and Ashlock's sense therefore means that the recognized taxon of descendants no longer shares derived patristic (apomorphic) characters. In the final analysis, these concepts again are based on phenetic similarity of descendants and ancestors, as further detailed in our critique of Simpson's definition of monophyly below. Yet another redefinition of these terms has been furnished by Nelson (1971).

Remane (1956) solves the monophyly-polyphyly problem more drastically by not accepting as natural any groups with characters which do not conform to his criteria of homology. Thus, by not recognizing polythetic taxa, he decreases the probability of a taxonomist's having to recognize polyphyletic groups. To admit the existence of polyphyly would be fatal to Remane's system, since he relies upon a closed circle of reasoning from monophyly to a natural system, to homologous structures, and back to monophyly.

Bigelow (1956) has pointed out that in all supposedly monophyletic classifications overall similarities and differences are usually not disregarded. Even in those cases where the ideal of a monophyletic classification could be attained, it often is disregarded in favor of a phenetic classification by supposedly phylogenetic

taxonomists. Bigelow feels that "if classification is to correspond with evolution, it must be based on the extent of overall difference, not on time."

The redefinition of monophyly (Simpson, 1960, 1961) is not free from ambiguity in its practical application: "Monophyly is the derivation of a taxon through one or more lineages (temporal succession of ancestral-descendant populations) from one immediately ancestral taxon of the same or lower rank" (Simpson, 1961, p. 124). We therefore need to know what is meant by the term taxon in each instance, and we need to decide the relative rank of the taxa. The difficulties were pointed out in *Principles of Numerical Taxonomy*. Simpson does not give clear criteria for deciding this rank. Mayr's (1969a, p. 75) brief discussion of the problem relies on distinctions between the phenotype and genotype of supposedly polyphyletic taxa. Since these distinctions can rarely be made it is hardly an operational criterion. In his criticism of Hennig's definition Mayr considers a rigidly cladistic interpretation of monophyly "contradicted by common sense" and implicitly modifies it by degree of phenetic diversity.

Let us return to the question of whether phylogenetic deductions can be made from phenetic similarities. If the resemblances are based on living organisms alone, we can only speculate on the phylogeny; to check our speculations we must have fossil evidence. Yet there are some conclusions that are more probable than others. We believe that these are as follows. (1) Phenetic clusters based on living organisms are more likely than not to be monophyletic *sensu* Hennig. Thus phenetically adjacent taxa represent phyletic "twigs," which usually originate from the same branch; in other words, overall convergence is unlikely. (2) In the absence of direct evidence our best estimate of the attributes of a common ancestor of a cluster must be derived from the properties of the cluster itself. In short, if we have no fossil evidence, the existing pattern is our best guide to the past history—though this may often be wrong. An argument similar to the first argument in favor of equal weighting of characters (Section 3.9) applies here: if we have no evidence that evolutionary rates differed, it is difficult to proceed further without assuming these to have been constant and equal in all the phyletic lines studied. If the reader thinks of a cross section through the top of a shrub with the vertical dimension representing time and the horizontal representing phenetic dissimilarity, he will have a ready, though somewhat inadequate, simile for the situation.

The two points mentioned can then be illustrated as follows. (1) Adjacent twigs will generally arise from the same branch. Admittedly it will be very difficult to detect substantial homoplastic similarity (overall convergence) near the tips of the branches, and this may have occurred quite commonly, together with some reticulate evolution due to the fusing of phyletic lines; yet gross degrees of overall convergence between the tips of the main branches will be very much less likely, and its improbability will increase with the taxonomic difference between the branches below the plane of section. (2) In the main, the branches from which the twigs arise will lie more or less directly below the twigs; but we will have no way of

telling whether the twigs arose almost vertically or whether they came off at a pronounced slant, for we can have no confidence that the twigs will fill the phenetic space evenly in the way in which the branches of an actual shrub do in order to obtain adequate sunlight.

The above deductions on the phylogeny (which are made from organisms belonging to one point in time) cannot give any estimate of the rates of evolutionary change, which may have differed in different phyletic lines. To estimate rates we must have data from several points in time.

These guide lines will hold only a simple form where one or two geological strata are represented. Where there are more than two, we will commonly be faced with the question: have rates of evolution been constant over the whole period? For example (see Figure 2-7), we may have an extant form **a** which is more similar to a slightly earlier form **b** than it is to a much earlier form **c**, though **b** and **c** are more similar. Did **a** then evolve recently from **b** (the more similar), or more slowly and directly from **c** (the less similar)? In the absence of other evidence, such as geographical isolation or a fuller fossil record, no certain decision can be taken even in such a simple case. It is very easy to find cases that are more complex than the example given above. One lineage, for example, may evolve rapidly and another slowly, and it may be impossible or implausible to draw a cladogram in which all the lineages evolve at the same rate throughout the period. A diagram of the phenetic relations (expressed in one or two dimensions) versus time will usually make clear the degree to which we can safely reconstruct the phyletic tree, and where we must indicate by dotted and queried lines our uncertainty as to the course of the descent.

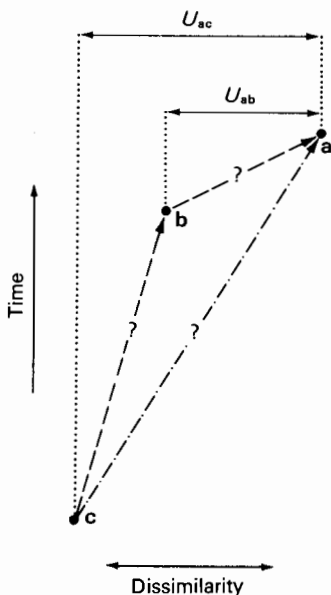


FIGURE 2-7

The time-rate problem. Organism **c** is the ancestor of both **a** and **b**. The phenetic dissimilarities between **a** and **b** and between **a** and **c** are indicated, respectively, by U_{ab} and U_{ac} . With only the data shown one cannot tell whether **a** arose directly from **c** or via **b**.

We have seen that one cannot derive evolutionary rates from similarity coefficients among recent forms. This is shown by the “pregroup-exgroup problem” discussed by Michener and Sokal (1957). Is it possible to distinguish whether an aberrant member of a cluster of forms was derived phyletically from one of the members of the cluster, or from the ancestral stem below the point at which the rest of the cluster arose (see Figure 2-8)? Michener and Sokal suggested that if the aberrant member, *x*, showed approximately the same resemblance to all the members of the cluster it was most likely “pregroup,” or derived from the common stem; see Figure 2-8,*a*. If, however, the similarity with one member of the cluster was much greater than the mean similarity with the cluster, then the aberrant member, *y*, was likely to be “exgroup,” or derived from the same stock as the member it most closely resembles—*d* in Figure 2-8,*b*. It is nevertheless possible to account for the observed resemblances in the figure by means of a number of cladistic schemes, if the evolutionary rates differ in the lineages.

Although the discussion of this section leads to the conclusion that cladistic relationships are difficult to estimate we should point out that the numerical cladists (e.g., Farris, Kluge, and Eckardt, 1970) have made considerable progress toward developing lines of reasoning which will determine cladograms with a fair degree of certainty (see Sections 6.3 and 6.4). These authors point out that even a weak correlation between phenetic similarity and cladistic relationship may provide useful information. In cases where the evolution of character states can be “safely” established, cladistic inferences are relatively simple. In other cases it is often necessary to undertake various types of phenetic analysis that are consistent with the assumptions of the cladogeny, using the evidence so obtained to develop hypotheses about character state evolution. These hypotheses in turn can be used to develop consistent cladograms based on a larger number of characters.

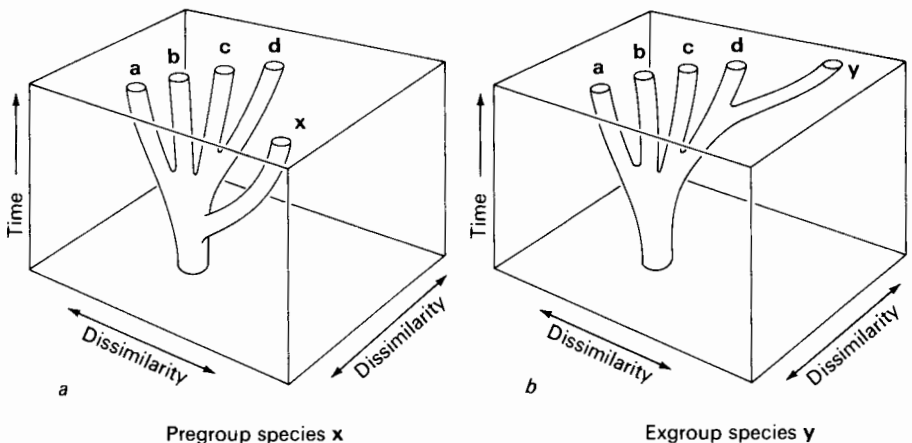


FIGURE 2-8

The pregroup-exgroup problem. For explanation, see text.

2.6 CHOICE OF A BASIS FOR CLASSIFICATION

We have seen that there are problems in both phenetic and cladistic approaches to the study of taxonomic relationships; which of these systems should be used by taxonomists in classifying nature? We may make our decisions at two levels of reasoning. (1) For which type of relationship are the problems of estimation less serious, hence more likely to be overcome? (2) Which type of relationship is inherently of greater interest and usefulness to systematists and biologists in general?

The present development of systematic theory and practice is such that the problems of estimating phenetic relationships (discussed in Section 2.4), though occasionally serious, are largely technical challenges that will be overcome before very long. In fact, discrepancies in taxonomic results attributable to these problems frequently yield useful information. By contrast, we do not have equally sanguine prospects for solving the problems of estimating cladistic sequences. Numerical cladistics may only provide a measure of compatibilities of characters, but not necessarily cladistic sequences. Numerous assumptions about ordering of characters and about evolutionary processes must be made to arrive at an estimate of cladistic relationships. Incisive mathematical techniques may be developed, which will solve some of these problems, yet the multiplicity of pathways for cladogenies may be too great to permit unequivocal decisions among them. More fundamentally, the question has recently been raised of the degree to which purportedly cladistic evidence is in fact phenetic evidence (e.g., Colless, 1967b). This point is discussed in greater detail in Section 6.3.

Turning to a discussion of the second question, we note that the difficulty with the use of the phylogenetic approach in systematics is that we cannot make use of phylogeny for classification, since in the vast majority of cases phylogenies are unknown and possibly unknowable. The theoretical principle of descent with modification—phylogenetics—is clearly responsible for the existence and structure of a natural system of classification; we may agree with Tschulok (1922) that the natural system can be considered as proof of the theory of evolution. However, since we have only an infinitesimal portion of phylogenetic history in the fossil record, it is almost impossible to establish natural taxa on a phylogenetic basis. Conversely, it is unsound to derive a definitive phylogeny from a tentative natural classification.

Again, we turn to Hennig for the most incisive analysis of the issues. Unfortunately, however, he derives erroneous conclusions from these. Aware of the limitations of the phylogenetic approach to taxonomy, he justifies such a procedure on the basis of four arguments (Hennig, 1966, pp. 22 ff.). The first is that the phylogenetic system is the most meaningful of all possible systems because all other types of classifications, such as ecological, zoogeographic, or morphological, can be derived and explained through the phylogenetic system. Indeed, none of the special classifications could occupy such a central and all-explanatory position

as does a phylogenetic system. The theory of evolution is the most adequate, most unitary, and indeed simplest hypothesis to which a great variety of biological phenomena—geographic distribution, physiological adaptation, morphological similarity, or biocoenotic complexity—can be related. Phylogeny can thus be seen as the central cause of much biology, yet it cannot be used for an explanatory concept, as phylogenies are not known in the vast majority of instances. Hence a phenetic classification, although it may not be able to explain the above-mentioned biological phenomena, is at least a self-sufficient, factual procedure and will in most cases be the best classification we can get.

Hennig's second reason for preferring a phylogenetic taxonomy has been negated by the development of numerical taxonomy. He states that phylogenetic relationships are at least in principle measurable, but that similarities are not. But the development and success of numerical phenetics has invalidated that argument. Hennig's third point is that although phenetic and cladistic classifications will be identical in many groups, to the extent that they differ, phenetic systems are less suitable as general reference systems for biological systematics. This supposed primacy of phylogenetic classifications is questionable and we discuss the issue at length below. Hennig's fourth point is that the problem of incongruence will make measurement of morphological similarity between taxa impossible. However, by adding characters of various life history stages and organ systems we can obtain reliable similarities between genomes in all stages of expression.

Even if we could make use of phylogeny to create classifications, we may still ask whether this is necessarily desirable. To do so would discard much important and interesting phenetic information. An allopolyploid might originate repeatedly, giving rise to phenetically identical new species each time. In some groups phyletic classifications might prove chaotic—for example in viruses and especially in bacteriophages. Work by Jacob and Wollman (1959), Lederberg (1960), and Cowie (1967) suggests that bacteriophages are being constantly derived from bacteria as genetic entities that acquire the properties of autonomy and parasitism. This work not only implies that bacteriophages which are identical (or almost so) may be polyphyletic, but, even more disconcerting, it suggests that they may be of composite origin. They are able to transfer genes from one bacterial form to another and to incorporate such genes into their own genomes; it is therefore possible that they derive some of their own genes from one host and some from another. Other viruses may be similar in these respects to bacteriophages, and this raises the disturbing possibility that, for example, an arthropod-borne virus may be part insect, part mammal, and part bird. Recent work by Subak-Sharpe et al. (1966) has shown that the pattern of small DNA viruses is very close to that of mammalian DNA, suggesting that these viruses may have arisen from mammalian DNA sequences. The only way to bring order into such a system is by a phenetic classification, such as those of Andrewes and Sneath (1958), Sneath (1962), Bellett (1969), or Gibbs (1969).

Even more disturbing to our notions of phylogenetic taxonomy is accumulating evidence (Loening and Ingle, 1967; Stutz and Noll, 1967) that chloroplasts of higher plants have evolved from symbiotic blue green algae. Thus there may be groups where the validity of a phylogenetic classification, even if possible, may be in serious doubt.

Cain and Harrison (1960b) pointed out that the principal disadvantage of phenetic classification is that some convergence or parallelisms may go unrecognized and that polyphyletic groups might be mistaken for monophyletic ones. However, the recognition of monophyly can scarcely be thought to be the only worthwhile or even the preeminent activity of systematists. In fact, the analysis of phenetic similarity and phenetic structure may be of as great or greater evolutionary interest in studies of adaptations of organisms to their environments, correlation with breeding structure, evolutionary rates (where time factors are known), analyses of ecological niches, and other concepts.

Sokal and Camin (1965) have discussed the issue of phenetic versus cladistic classifications at length and much of what follows is directly based on their account. Since phenetic, cladistic, and chronistic aspects are necessary to understand the systematics of organisms, Sokal and Camin feel that systematics as a whole must be based on all of these considerations. Simpson (1961) has called phenetic taxonomy "shallow and incomplete." From the viewpoint of the entire field of systematics this is correct and would be equally true of a phyletic taxonomy based only on the tracing of cladistic pathways. Systematists taking an extreme position on either phenetics or cladistics are like the blind men trying to describe an elephant in the legendary folk tale (made memorable for Americans through a poem by J. G. Saxe), each man describing the whole elephant on the basis of the part he can touch. The difficulty with a composite approach to phylogenetic taxonomy, as espoused by Simpson, is that no operational methods exist for a synthesis of the various approaches. Basing taxonomy on all three approaches requires art or compromise, both of which are inadmissible as bases for a precise science. The acrimonious debate as to which system is better is thus of little profit. The question is for what purpose a classification is to be established.

An important consideration in any discussion of problems of systematics and classifications is recognition of the duality of systematics (Sokal, 1964a). One branch of systematics, classification, has the relatively unexciting but supremely important function of ordering nature into a practical and generally useful system. This can best be done on the basis of a phenetic classification, as will be shown below. Systematics in the wide sense also has the more challenging task of understanding the mechanisms which have brought about this order. These two aspects of systematics are separable and should be separated. The reluctance of conventional phylogenetic taxonomists to recognize this separation may be based on a conscious or subconscious feeling that classification by itself is an unimportant enterprise of low caliber and not worth doing. We think otherwise and are

supported in this view by many eminent systematists (e.g., Mayr, 1968). In fact, taxonomists often justify their activities by claiming that other biologists would be unable to work unless they knew what organisms they were working with. For this entirely legitimate reason a stable and consistent system of classification and nomenclature must be developed and maintained. By not allowing for the separation between classification and what Michener (1963) has called the "explanatory element," conventional systematists perpetuate a system which is inherently unstable and hypothetical by the very nature of its operations.

Because phenetic classifications require only description, they are possible for all groups and are more likely to be obtained as a first stage in the taxonomic process. This trend is likely to continue and be reinforced by the continuing development of automatic data-recording equipment of various types. On the other hand, since cladistic classifications require more knowledge about the organisms and the direction evolution is taking, they must necessarily follow phenetic clustering. If a uniform system of classification is to be established, it might therefore be argued that it should be done on a phenetic basis, with a complementary cladistic classification added whenever possible.

From the viewpoint of evolutionary study, both phenetic and cladistic classifications are of great interest, leading to an understanding of evolutionary principles. From the point of view of biology in general, however, it is probably of more interest to describe the existing overall similarity of the organisms, rather than their cladistic affiliations. This argument may be denied by those whose interests in biology largely concern evolutionary rates and lineages. However, not all biologists have such an orientation. The outlook of many physiologists, biochemists, behaviorists, and others is purely functional. The primary use of a system of classification for these biologists is to group organisms sharing as many properties as possible so that predictions can be made from one member of a taxon to others. Thus a biochemist finding a new substance in the blood of a species of *Rana* would be tempted to look at other members of this genus for further occurrences of this substance. Although such a phenomenon undoubtedly has an evolutionary basis (patristic similarity, see Section 2.4), as far as this application is concerned the process of prediction is based on the purely phenetic properties of the taxon. Similarly we use taxonomy as an efficient device for summarizing detailed information on many taxonomic units and for predicting properties of members of taxa. Thus, when told that a new species of the genus *Rana* has been described, we can immediately make many predictions about its appearance without ever having seen it. This follows from the naturalness of the genus *Rana*, in the sense in which we use the concept in this book.

In most cases where primitive patristic elements constitute an appreciable portion of the phenetic similarity, the cladistic and phenetic classifications will largely coincide. However, in the possibly rare cases where this is not so, the general superiority of a phenetic over a cladistic classification becomes even more evident.

This is illustrated by the admittedly implausible case of Figure 2-9 in which the evolution of line *x* of cluster **A** into the phenetic region of cluster **B** has made *x* resemble the members of cluster **B** to such a degree that a phenetic classification would include it in **B**, but somewhat closer to cluster **A** than most members of **B**. Although the divergence of *x* from cluster **A** is of great evolutionary interest, the overall similarity of *x* to the members of cluster **B** is more generally useful. Predictions of the characters of *x* will be more successful when based on characters known for taxon **B** than for those based on taxon **A**, because of *x*'s greater overall similarity with **B**. But it is also true that additional knowledge about the cladistic relationship of *x* to taxon **A** will further aid predictions about characters of *x*. Thus we can see, even from this simple example, that phenetic as well as cladistic relations are important.

In this connection we should mention an often misunderstood point. Natural classifications in Gilmour's sense are not necessarily cladistic (and vice versa). In the example in Section 2.2 organisms *e* and *f* might be phylogenetically convergent on organisms *a-d*, giving **A'** and **B'** as the cladistic groups. Yet groups **A'** and **B'** would not contain the most information in the sense meant by Gilmour, even if the additional information about their origin was added as a sixth character, that is, "belonging to clade **A'** versus clade **B'**." For then Groups **A** and **B** would have respectively five and five statements at the 80 percent level, but groups **A'** and **B'** would have only five and two statements, respectively. Colless (1969d) points out that if particular cladistic groups are preferred on the grounds that they are maximally predictive, this means that they must also be phenetic groups.

Decisions on which system of classification to adopt are affected by the method of representation of the taxonomic structure that is employed. Verbal or written descriptions of relations among organisms have proved quite inadequate. For this

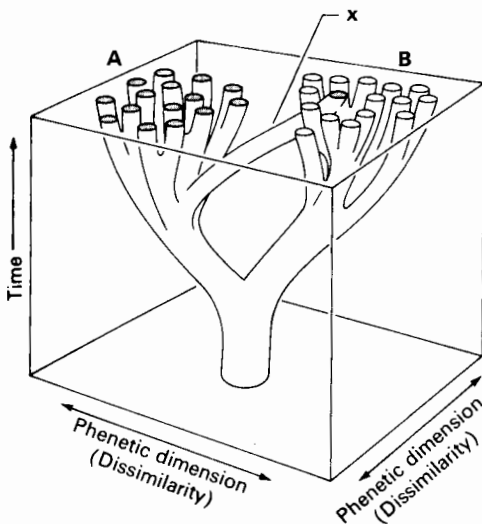


FIGURE 2-9

Phylogenetic tree to illustrate possible incongruence between phenetic and cladistic relationships. Phenetic relationships are shown in two dimensions, the third dimension is time. Cladistically, *x* belongs to the monophyletic taxon **A**, but phenetically it would be considered part of taxon **B** because of extensive parallelism. [Modified from Sokal and Camin (1965).]

reason a variety of mnemonic and didactic aids have been developed, most of them graphic. These are largely different forms of trees of relationships (or phylogenetic trees). Mayr, Linsley, and Usinger (1953) have called these drawings *dendrograms*, which seems a suitable term without any implication about the nature of the relationship. The terms *phenogram* and *cladogram* (Mayr, 1965; Camin and Sokal, 1965) have come into general use to define dendrograms representing phenetic and cladistic relationships, respectively.

A dendritic description of the taxonomic system has much to recommend it and seems in many ways to be the "natural" way of illustrating relationships and descent. A vague, general agreement on the interpretation of diagrams of relationship exists among taxonomists, yet when a given diagram is subjected to detailed, critical scrutiny we rarely find consistency of meaning within it. The interpretation of the basic facts that a diagram offers is likely to be based on varying degrees of certainty in different parts of the tree. No generally accepted conventions for constructing such diagrams exist; hence, seemingly similar diagrams may have quite different meanings, which are often not clearly enunciated by the author of the tree. In an illuminating discussion on the different components that can be included in phylogenetic relationships, Cain and Harrison (1960b) have shown that often an author himself has not a clear idea of the meaning of a diagram of relationships he presents.

The following symbolisms have been used most frequently.

1. The vertical axis (or radius in circular dendrograms)—to indicate time, either in absolute units or in relative evolutionary ones (most frequently unspecified).

2. Furcations—to indicate branches in the phyletic sequence, and so to show the relationships between the forms based on the lineages alone (not considering their phenetic similarities); that is, the cladistic relations.

3. Location and relative position of tips of branches with respect to each other—to indicate phenetic relationships.

4. Location of furcation along a vertical axis (which now designates resemblance)—to indicate closeness of relationships between taxa represented by stems issuing from the furcation. Symbolisms 3 and 4 are often used in combination to indicate what part of the relationships is due to convergence and what part is due to patristic similarity.

5. Levels of tips along a vertical axis (or along the radius of a circular dendrogram)—to indicate whether the forms are extant or extinct, and also to give some estimate of the time scale of the extinction.

6. Levels of tips along a vertical axis (or along the radius of a circular dendrogram)—to indicate degree of perfection or complexity of form. This convention, related to the *scala naturae* of earlier science, is largely out of fashion, although some of its ideas and its vocabulary are still employed occasionally. Thus Rensch (1947) uses the term *Höherentwicklung* (anagenesis).

7. The angle between stems—to represent velocity of phenetic differentiation.

8. Thickness of stems—to represent abundance at a given point in time. Abundance is usually measured by the number of species or taxa contained within the stem, but occasionally represents the number of organisms supposedly extant.

It is easily seen that 1, 2, and 5 can be combined into a single diagram. Unless the rate of evolution has been constant, 4 cannot also be included. It is generally impossible to represent phenetic relationships on a two-dimensional graph; hence 3 is bound to be a distorted representation. The use of the angle to indicate velocity of evolutionary change, 7, is never very successful except in the simplest diagrams. Abundance, 8, can usually be added to most diagrams, although the results are often not very esthetic. The basic difficulty is the graphic representation of phenetic resemblances and phenetic change. These are multidimensional relationships and cannot satisfactorily be compressed into a two-dimensional diagram. The difficulties of representation have led to other schemes of presenting the relationships, but these have not been too successful. Increasing familiarity with multivariate procedures has resulted in new attacks on problems of representing phenetic similarity among taxonomic units (see Section 5.9 for an extended discussion), and it has been realized that earlier attempts to force these into hierarchic relationships by means of phenograms have not always resulted in the most satisfying classifications.

Hennig (1966) states that for similarities among organisms to be recognized as valid expressions of taxonomic relationships, they must permit the taxa to be arranged in a hierarchic system. We disagree with this point of view for the following reasons. Hybridization may well have occurred more often than we realize in animals (and is of course well known and frequent in plants). It may therefore be impossible to make hierarchic relations in some groups because of extreme reticulate evolution. Possibly in consequence, morphological similarity relationships can sometimes be represented better in nonhierarchic ways (see Section 5.3) than by the traditional hierarchic system. Hennig (1966, p. 21) is therefore incorrect in stating that relationships of morphological similarity between species can best be represented in a hierarchic system. It should, however, be noted that reticulate phenetic relationships can often be approximated satisfactorily by a hierarchic, dendritic relation. Thus, for example, the reticulate relationships among six species of the genus *Cyrnus* illustrated by Hennig (1966, Figure 20) can be shown to yield a satisfactorily hierarchic structure.

Hennig feels that it is fallacious to assume similarity relationships between living organisms to be primary, but that genealogical relationships are primary. This seems to us to be similar to the argument that either the male or the female is primary in reproduction. The critical issue is not which is the more important relationship, but which is the relationship that we *can* investigate and from which we may derive useful principles. On these counts phenetics clearly qualifies. The only possible way of transmitting the various types of information listed in points

1 through 8 is by three separate graphs for (a) time and branches (cladistic relationships), possibly combining symbolisms 1, 2, 5, 7, and 8; (b) phenetic relationships between junctions of stems only, as customarily employed in dendrograms in numerical taxonomy—symbolism 4; and (c) complexity of form or organization—symbolism 6. Phenetic relationships among tips of branches (symbolism 3) usually cannot be represented adequately even in a two- or three-dimensional space. Proper representation would require two-dimensional cross sections through the hyperspace that is necessary to represent such relations. We shall consider the mechanics of such a presentation in greater detail in Section 5.9.

As regards aspect 2—phyletic sequence—it is obvious that a diagram can be constructed only if phylogenetic evidence can be obtained from fossils or in some other reliable fashion. We have already pointed out the dangers inherent in inferring phylogenies from Recent organisms. The sequences in phyletic lines are often much more uncertain than authors wish to admit. Some authors indicate probable descent by dotted lines. If there are many such dotted sections, the chances of the diagram being substantially and seriously misleading may be very high indeed. Unfortunately there seems to be no study on this point to tell how misleading earlier phylogenies have been when compared with later detailed and convincing fossil evidence. Such a study might be illuminating. It is true that many authors, quite properly, disavow any phylogenetic significance for their diagrams and caution readers against considering them to be in any way reflections of evolutionary history. We ourselves frequently follow conventional practice in arranging taxa by a system of hierarchic, nested categories which roughly give an indication of point 4. This and other forms of representation are taken up in Section 5.9.

In summary, for systematics to be thoroughly understood, phenetics, cladistics, and chronistics must all be considered. For classification, that is, for the formal description and cataloging of organized nature, phenetics appears preferable.

2.7 TAXONOMIC RANK

Criteria for Taxonomic Rank. Phenetic as well as genetic criteria are commonly used; at and below the level of biological species these criteria may be in conflict, however. In the absence of data on breeding and in all apomictic groups (which together include the great majority of practical applications in systematics), the criteria for species are based on the phenetic similarity between the individuals and on phenotypic gaps (Michener, 1970; Sokal and Crovello, 1970). The rank of higher categories must perforce depend on phenetic criteria. The intrusion of an entirely different criterion for taxonomic rank in those few situations where genetic or phyletic relations are known with certainty, seems to us to be a needless source of confusion. Alternative terminologies have been suggested for genetic and also ecological entities (especially in botany, where these problems are most acute; compare Gilmour and Gregor, 1939; Camp and Gilly, 1943; Gilmour and Heslop-

Harrison, 1954); these terms have not been widely used. For reasons of clarity it is desirable that the meaning in which taxonomic rank is used should be specified. In this book it will be used in the sense of phenetic rank, unless otherwise indicated. A detailed discussion of the species concept is deferred to Section 7.1. Jardine (1969b), in erecting his logical system for taxonomy (see Section 2.8), finds little justification for considering any one category in a natural biological classification more or less real than any other category except in the sense that some diagnostic criteria may be more or less reliable than others, as for example gene flow in studies of putatively specific populations. This position is retained in his subsequent work (Jardine and Sibson, 1971).

It is undesirable for the rank of a group to be affected by the number of contained subgroups. There is a modern tendency to make each family contain only a few genera and each genus only a few species; in some works most genera are monotypic. The rank should be based on relative degree of similarity alone. In our view it is better to introduce new rank categories (such as subfamily or superclass) than to use the number of contained subgroups as an arbiter of rank.

There are as yet no criteria for any absolute measure of taxonomic rank. We do not know how to decide whether a family of birds, for example, is equivalent phenetically to a family of insects. The hope has been expressed by one of us (Sneath in Heywood and McNeill, 1964, p. 159) that protein sequence studies might lead to such criteria, but it is evident that this was overly sanguine; different types of protein evidently show widely different phenetic differences at the same taxonomic ranks. Recent work on DNA (e.g., Park and De Ley, 1967; Dessauer, 1969) seems also fraught with problems. The difficulties in using the age of clades as a measure of taxonomic rank are well discussed by Hennig (1966, p. 183).

Limits of Taxa. The limits of taxa have sometimes been suggested as criteria for taxonomic rank. Limits can be considered from two points of view. One can trace the change of taxa with time, looking at the phylogenetic tree in its entirety. This is what Simpson (1945) has called vertical classification. One can also look at a cross section of the tree and obtain the relationships among taxa at a given point in time (horizontal classification, Simpson, 1945). With vertical classification (which can only be practiced on fossil material) it is obvious that when one phyletic lineage evolves into a new form there can be no sharp division between the ancestral and the descendant species, other than an arbitrary one, except in the case of allopolyploids and other forms of hybrid origins. The accidents of discovery of fossil forms inevitably affect classificatory decisions, since the divisions will at first be placed where there are gaps in the fossil record. As the gaps are closed by new discoveries, the most common practice is to choose for the dividing line some prominent, but commonly arbitrary, evolutionary step—for example, the change in jaw structure in the evolution of reptiles into mammals. So long as the arbitrariness is clearly realized, these methods are unobjectionable and are matters of

convenience. The choice of such an arbitrary step is not without some danger, however, for it may lead to incongruous situations.

A better plan, commonly advocated in paleontology when a relatively full fossil record is known, is to place the divisions at places where abrupt changes in the rate or direction of evolution make for rational and convenient groups or where phyletic lines branch. The demerit of this course is that the divisions then come at those very parts of the lineages which are of special interest for students of evolution. Nevertheless, the bulk of the total material will be grouped in a convenient way.

To turn to horizontal classification, difficulties will arise if a phenetic group is not identical with a phyletic group. For example, the appearance of a sterility barrier will at once divide a normal genetic species into two sibling species (from the point of view of the "biological" species definition). Yet for many generations (until the two sibling species have accumulated sufficient genetic differences in the course of their independent evolution) they may remain one single phenetic group, because the differences that cause the sterility barrier (plus the few other accumulated differences) will be insignificant in comparison with the many variable attributes of other kinds that the two sibling species will share. The sibling species will therefore constitute a single natural taxon, in the sense in which we use the term. Indeed, our point has apparently been misunderstood by Mayr (1965) who criticizes numerical taxonomy because it would in such a situation group together the sibling species. We must emphasize that this is precisely what it is intended to do; a phenetic method should give phenetic groups.

A point that has had little attention, although Simpson (1961) and Michener (1963, 1970, p. 23) refer to it, is to what extent taxonomic rank is represented by variability within a taxon, the distances between the centers of taxa, or by the size of the gap between them. The last may, of course, be quite small if the taxa almost overlap. Nor is it clear whether some compromise is desirable. This problem appears to underlie the difficulties of Gregg's Paradox (see Section 2.8) where taxonomic ranks are defined in terms of set theory (Gregg, 1954). It is clear that if the rank of a taxon is measured by its internal variability one may have the paradoxical situation where the only species of an order, and consequently the order itself, has very little variability. It also appears undesirable to base rank primarily on the size of gaps between taxa; rare intermediate forms may then become critical in recognizing rank. Because of the logical relationships between the phenetic distances involved in these criteria, the question is best discussed in detail in connection with the topic of patterns in phenetic space in Section 5.11.

Taxa of low rank may be difficult to define and to arrange hierarchically. Such groups would appear in a numerical taxonomic study as contiguous and indistinct clusters of individuals. Methods for studying this are discussed in Chapter 7.

The Hierarchy of Characters. This refers to the claim—fallacious, and fortunately on the way out—that one can lay down a priori rules as to which kinds of characters

separate species, which kinds separate genera, which kinds separate families, etc. This is the antithesis of the a posteriori method of discovery, which *finds* those features that do in fact separate the previously recognized natural taxa, and is, we maintain, the correct procedure. The concept of a hierarchy of characters is now mainly of historical interest, being specially connected with the French systematists (the de Jussieu, Cuvier, and de Candolle), and is often termed the principle of subordination of characters. The rules were avowedly based on the fancied importance of different organ systems. Yet they were qualified by so many exceptions, as well as by appeals to the constancy of characters within taxa (a presupposition that the taxa are already established), that it is clear they were never workable rules (Cain, 1958). We believe that no such hierarchies can be made a priori. It would indeed be curious if evolution, which is responsible for the natural hierarchy, should be so obliging as to operate only on certain classes of characters at specified taxonomic tanks. Some taxonomists prefer to base their classification on what they suppose to be nonadaptive characters. Population geneticists are currently debating vigorously whether any characters can be considered nonadaptive. Even if such characters exist, it would be almost impossible for taxonomists to recognize them as such. The converse view—that taxa should be based on adaptive characters (Inger, 1958)—is quite impracticable, as has been shown by Sokal (1959).

2.8 DESIRABLE PROPERTIES OF A TAXONOMIC SYSTEM

A taxonomic system will reflect the purposes for which it is constructed, and we will therefore need to specify these in some detail. It is a measure of the current re-evaluation of taxonomy, with its consequent uncertainties, that different taxonomic textbooks (e.g., Simpson, 1961; Davis and Heywood, 1963; Hennig, 1966; Blackwelder, 1967a; Mayr, 1969a; Jardine and Sibson, 1971) all give different emphasis to the details of these purposes. It may be that certain purposes—and hence, properties—are incompatible, and the taxonomist must then decide which to prefer, or whether to look for alternative purposes that represent some sort of compromise. Deutsch (1966) has made a first attempt at providing operational and quantitative definitions of some ways for evaluating taxonomic schemes. He lists the following as characteristics of a taxonomic scheme:

- organizing power
- predictive range
- structural complexity
- functional complexity
- performance
- novelty of performance
- structural novelty
- functional novelty

predictive accuracy
 social costs (inverse of "economy")
 net change in accuracy
 net change in cost
 originality
 fruitfulness
 self-transcendence

For operational definitions of these characteristics, some of which are clearly applicable to biological classifications, see the original reference and a similar discussion by Fisher (1969).

In the past, numerical taxonomists have given much attention to imitating existing taxonomic practices. This may, however, not be the most profitable approach. We may be able in the future to develop quite new methods that are not intended for the traditional purposes of taxonomy, but which have value for biology in general as well as for systematics in particular. This is still in the future; the present is a time of rapid change of taxonomic ideas, and it is difficult to foresee the detailed needs of even the immediate future of systematics. The following discussion is therefore written largely from the traditional point of view; that is, we discuss the properties of classificatory systems that taxonomists consider desirable today.

Desirable properties and purposes of classifications are linked together in a way that makes them difficult to discuss individually; most have repercussions on many others. We shall divide them roughly into three sections: (1) those concerned with "naturalness" of taxa; (2) those concerned with ease of manipulation; and (3) those that facilitate retrieval of past information.

Naturalness. The basis of natural classifications has been discussed in Section 2.2, and it is clear that natural classifications (in some sense of the word classification) are what are principally required. There is of course a place for special classifications for special and restricted purposes; these pose their own problems, but these are often peculiar to the immediate study. In general, consistency and accuracy are the main needs.

In Section 2.2 reasons were given for preferring a natural system in Gilmour's sense, and in Section 2.6 we have explained why we believe that phenetic rather than phyletic groups are desirable.

Natural and polythetic phenetic groups are thus desirable, and we have seen that these possess high information content, and also have a high predictive value for characters of new organisms. For example, Watson, Williams, and Lance (1966) have emphasized the importance of this, saying "those who need a new source of an unusual plant product turn first to the taxonomic relatives of species known to produce it." The requirement of naturalness also implies that some criterion of optimality is used in the class-making steps; this may be at a simple level, as in the

example given in Section 2.2, or may be more sophisticated, depending largely upon what information is to be assumed about the system—whether, for example, size factors are to be ignored so that only information about shape is included. A more extensive discussion of optimality criteria is given in Section 5.10. In view of the proliferation (both of statistics and also of concepts) relating to different components of phenetic resemblance, it is particularly important that the maker of a classification should have a clear idea of what he wants, and he should indicate what components are used as the basis for his classification.

Since the time of Linnaeus taxonomists have thought of the system of nature as nested and mutually exclusive. Early attempts to define nonoverlapping, taxonomic hierarchies in terms of symbolic logic were made by Woodger (1937, 1951, 1952) and continued in an important book by Gregg (1954). In Gregg's study, a taxonomic system was defined as a sequential partition of a set of taxonomic units into disjoint subsets, each representing a taxon with its rank given by the number of partitioning steps it has undergone. The definition of taxa in this context led to a difficulty that has become known in the literature as "Gregg's Paradox." This applies to monotypic taxa that conventionally differ in rank, yet would not do so under Gregg's model. Several authors (Sklar, 1964; Van Valen, 1964; Buck and Hull, 1966; Farris, 1967b; Gregg, 1967, 1968) have attempted to overcome this difficulty.

The most recent contribution has been a new logical model for taxonomy developed by Jardine, Jardine, and Sibson (1967), extended by Jardine and Sibson (1968a), and reviewed in semitechnical form in Jardine (1969b). Using set-theoretical definitions, Jardine arrived at a definition of a taxon as an equivalence class, that is, a set of objects related by an appropriately defined equivalence relation. Each taxon Y in a nonoverlapping taxonomic hierarchy is defined as an ordered pair of terms, the first member being *Ext Y*. This is short for *extension Y* and implies the entire set of objects or taxa that share the particular relationship implied by the equivalence relation at a specified rank in the hierarchy. The second member in the definition of a taxon is its *rank*, defined as $J(\mathbf{a}, \mathbf{b})$, the lowest ordinal rank of the taxon of whose extension \mathbf{a} , \mathbf{b} are members. Since by Jardine's definition one can distinguish taxa with identical extensions but different ranks, monotypic taxa do not encounter the difficulty of Gregg's Paradox. The basic units in Jardine's system are what he calls basic taxa and what we would call OTU's (see Section 3.1). These may be individuals but are more likely descriptive vectors or matrices representing populations. We shall not enter upon a detailed explanation and derivation of the system established by Jardine and his associates here, but refer the reader to the sources already quoted. It is important to note, however, that Jardine's nonoverlapping hierarchic system leads to six theorems, which can be derived from his model and are generally accepted statements about taxonomic hierarchies. Fundamentally they concern the "nestedness" of the hierarchy, except for his last theorem, which shows that a nonoverlapping hierarchy must obey the ultrametric inequality (see Section 4.2).

Given a natural nonoverlapping taxonomic hierarchy, it can be represented in the form of a dendrogram because of its ultrametric property (see Section 4.2). However, as pointed out by numerous authors (Jardine, 1969b; Farris, 1969a; Hartigan, 1967) dissimilarity matrices are generally not ultrametrics and hence cannot be mapped by a one-to-one relationship into a dendrogram. But when the similarity or dissimilarity coefficient is not even a metric, serious distortion of the taxonomic relationships and “reversals” frequently occur when the results are represented in dendritic form (see Sections 4.2, 5.5). The degree to which such dendrograms and their corresponding classifications distort the similarity relationships among the assemblage of objects represented by the classification is estimated by various measures of stress and distortion discussed in Section 5.10.

An important, if somewhat poorly defined, criterion is concordance between a classification and other facts about the organisms. This applies at two levels, the logical and the technical. At the logical level it is an invariable test of goodness in all branches of statistics because this is our only way of checking the validity of any hypothesis; if challenged we must resort to additional evidence. There is, then, a good reason for saying that an important test of a classification is whether it agrees with another classification of the organisms based on additional data, or in other words, exhibits stability upon addition of new information. There are some problems in setting up completely logical and self-consistent tests of this kind (Sneath, 1967b), notably problems about the validity of the new characters and organisms to be employed. Also, it is true that lack of concordance does not necessarily condemn a classification, because the incongruence between classifications based on different stages of the life cycle may have a biological explanation. This is also to some extent a side issue, for if our aim (as it generally is) is to classify organisms (life cycles) rather than simply specimens (life stages) we should incorporate as much information as possible from all life stages and all organs and character-systems. The amount of information from each life stage (or organ) will thus depend on the amount of information known about it, and this would seem to be the best way of obtaining a representative set of information about the organism *as we know it*.

At a more technical level, we would rightly be suspicious of a classification that violently disagreed with certain kinds of technical evidence, with serology or DNA pairing studies, for example (provided that the evidence was beyond question). These external criteria must be viewed with common sense; they must be both biologically sound and also highly informative in the sense of containing much information. Thus there are general grounds for believing that serology, DNA studies, and protein sequences reflect a great deal of genetic information. In contradistinction to this, the presence of a simple chemical compound would be a poor court to appeal to; one such substance would be quite unacceptable, although a hundred might provide a good check upon a given classification.

It may be noted that both of the last two paragraphs are tests of badness of taxonomies rather than of their goodness. Concordance does not ensure that the

classification is good, but discordance implies that it is bad. Nevertheless a number of such external criteria which all agree will make for a near certainty that the classification is valid. We do not see that there is much difference here between testing the validity of taxonomies and any other scientific phenomena.

As much information as possible is desirable for constructing taxonomies, for three reasons: to allow enough power to the method (e.g., a single presence-absence feature would obviously only permit the recognition of two taxa, the "haves" and the "have-nots"), to give a representative cross-section of attributes of the organisms, and to reduce chance and sampling error to a minimum.

Ease of Manipulation. Most if not all of the above points are to a large extent compatible. We now turn to some that may frequently not be readily compatible with the above requirements. Criteria that relate to the working ease of a taxonomy include a hierarchic arrangement if it is possible (it may not always be), and the ability to produce good taxonomic keys (i.e., discrimination between taxa should be as easy as possible). These criteria may be in conflict with those of naturalness described above. Thus it may be unwise to force a hierarchic arrangement onto material that does not possess the structure necessary for a hierarchy; on the other hand the memory-saving properties of hierarchic systems are of considerable practical importance. Discriminatory requirements may also conflict. It is too readily assumed that taxonomies should be based on measurements in what may be loosely termed "units of discriminatory power," at the expense of a faithful representation of phenetic resemblances and of the information content of the taxa. We would hold that discrimination is a rather special purpose of taxonomies, and as a general rule phenetic resemblances are better suited to most taxonomic work.

Williams and Dale (1965) have pointed out that classifications can either be probabilistic or nonprobabilistic. The former require a null hypothesis, and frequently no suitable test of such a hypothesis may be available. There are, of course, advantages in probabilistic models when these are feasible, but this is by no means always the case. Again, we would view this as a desirable additional property when possible, but not an overriding one.

Information Retrieval. We believe that the retrieval of information, although often placed high on the list of priorities (for example, one often hears the dictum that "a plant's name is the key to its literature"), is in fact subsidiary to questions of naturalness. If groups are indeed natural, the main requirement for easy information retrieval is a suitable nomenclature. This therefore becomes a subsidiary desideratum of a taxonomic system, important, but not as important as natural taxa, and we believe that in the future ready adaptability to electronic data processing will become an important element in any nomenclature.