CHAPTER 6

THEORY AND PRACTICE OF BIOLOGICAL CLASSIFICATION

Objects in nature seem to fall into groups or classes, such as birds, butterflies, trees, grasses, bodies of water, and stones, with the members of each group having one key attribute or a number of attributes in common. The traditional definition of *classification* therefore is *the grouping of objects into classes owing to their shared possession of attributes.* Any ordering of the vast diversity of nature, or of any part of it, into such groups is called a classification. Communication would be difficult if such classes were not distinguished and named. Used in all human activities, classifications also exist without benefit of language. Animals prove by their reactions that they classify objects in their environment as food or nonfood, competitors or potential mates, enemies or prey. Michener (1970:5) put it this way: "A bird that classifies a cat as one of the rabbits is not likely to maximize its reproductivity."

SOME BASIC PRINCIPLES OF CLASSIFICATION

In view of the ubiquity of classifications and the corresponding need for an understanding of the basic principles of classifying, one is perplexed by the neglect of this subject in the philosophical literature. Some philosophers, such as Nagel in *The Structure of Science* (1961), do not even mention the problems of classification. Others, from ancient (Aristotle) to recent times (Whewell, Mill, and Jevons to Hempel and Popper), have

referred to classification, but they have dealt with it typologically and often on the basis of physicalist principles. Most philosophers have failed to see the difference between classification and identification schemes and have relied on logical division (downward classification), a method which leads to identification schemes but not to genuine classifications. Almost all of them have assumed that there is no difference between the classification of inanimate objects and that of organisms.

As we shall see below, a classification of living organisms must take into account certain special constraints (evolutionary history), yet at the same time it should not violate certain elementary rules of classification that are equally applicable to both organisms and inanimate objects. Curiously, we have been unable to find in the philosophical literature a well-articulated list of such rules.

General Rules of Classification

The attempt is made here to fill this gap in the literature by supplying a set of general rules of classification. These elementary rules, here tentatively proposed, are the ones that are in daily use in arranging books in a library, goods in a warehouse, or character traits in psychology. Additional rules, which are supplementary, but not contradictory, are needed for the classification of organisms.

1 Items that are to be classified are assembled into classes that are made as homogeneous as possible.

2 An individual item is included in that class with the members of which it shares the greatest number of attributes.

3 A separate class is established for any item that is too different to be included in one of the previously established classes.

4 The degree of difference among the classes is expressed by arranging them in a hierarchy of nested sets. Each categorical level in the hierarchy expresses a certain level of distinctness.

Surely there must be more general rules than these four, but they will suffice for this provisional listing.

Additional Rules for the Classification of Organisms

Most philosophers and even a number of taxonomists, among whom Gilmour (1940, 1961) and his school have been the most prominent, have chosen not to make a distinction between classifying organisms and classifying inanimate objects. In either case, they state, that classification would be the most natural that served the largest number of purposes (for critiques, see Simpson 1961:25 and Farris 1979*b*:497).

Eventually taxonomists, Darwin most decisively, realized that it is neither useful nor legitimate to classify objects or phenomena strictly on the basis of some defining quality (essence) or in any other arbitrary manner when the grouping of these objects is actually the result of history or another cause (Mayr 1982*a*:238). Such forms of causation exert constraints that severely limit the number of possible meaningful classifications. For instance, classifications of human diseases by means of such arbitrary and artificial criteria as "quick recovery versus slow or no recovery," "with fever or without fever," and "with pain or without pain" were found by physicians not to be very helpful. A far more appropriate system of classification sorts diseases into those *caused* by bacteria or viruses, malignancies, degenerative changes, and inheritance; that is, it sorts them by causal agents. Darwin was the first to see clearly and state emphatically that the grouping of organisms is a result of the common descent of the members of a group (taxon). In order to be "natural," a classification of organisms therefore has to reflect descent. This theoretical basis of all natural biological classifications is a powerful constraint and refutes the claim that the same principles of classification are applicable to organisms and to inanimate objects. From these considerations one arrives at the following definition: *A biological classification is the ordered grouping of organisms according to their similarities and consistent with their inferred descent.*

It is thus evident that there are two independent sets of criteria responsible for the grouping—*similarity* and *sameness of causation* (lines of descent)—on the basis of which organisms can be assembled into taxa. As we shall see, these two sets of criteria are frequently in conflict, and their application is controversial. In fact, the most important difference among the three major current schools of taxonomy lies in the relative weight they give to these two criteria. These differences are the subject of this and the next five chapters.

Special Classifications

Classifications that attempt to reflect a similarity and/or relationship are based on numerous ("greatest possible number of") characters. For practical purposes other, special classifications based on single characteristics are sometimes needed, for instance, diploid versus polyploid plants, annual versus perennial herbs, and inedible (or poisonous) versus edible mushrooms. A limnologist may divide plankton into autotrophs, herbivores, and predators. Plants are traditionally subdivided into trees, shrubs, herbs, and grasses, and this grouping is still useful in some areas of ecology. Such special-purpose classifications have a low information content and usually cannot be used for broader generalizations.

One must remember that the taxonomist classifies populations at all levels, not characters. Characters are only the means by which taxa are recognized. See Chapter 9 for the dissenting views of pattern cladists.

IDENTIFICATION

There is a fundamental difference between a classification and an *identification* scheme. A classification orders a diversity of items into groups or taxa on the basis of principles and criteria that are discussed below. An identification scheme or key permits the placement of an unidentified object (specimen) in one of these taxa. In identification one uses a few characters, ideally a single diagnostic one, that throw a given specimen into one line or another of a key. Computer identification methods work somewhat differently but also focus on the specimen that is to be identified. The procedure of identification is based on deductive reasoning.

All so-called field guides for the recognition of birds, butterflies, flowers, or other organisms are based on identification schemes. Identification keys are also valuable components of monographs and revisions. However, the preparation of these keys must be preceded by a careful taxonomic analysis. Identification keys and classifications serve very different purposes, yet the quality of an identification key depends largely on the quality of the classifications on which it is based. It is wrong to say "I classified a specimen" when one means "I identified a specimen."

A classification is a filing system, and the unique name of a species or higher taxon is like the index number of a file. It is the key to the entire literature on that taxon. Many of the developments in taxonomy in the last 200 years have led to a clearer separation of the procedures of identification and those of classification. Prior to Linnaeus, virtually all so-called classifications were actually identification schemes; so were the artificial downward classifications of Linnaeus (Chapter 1). They were based on a method of logical division in which the organic world as a whole was considered a summum genus, which at the first step of the identification was divided into two "species." With the last of a series of dichotomous divisions, one arrived at the species of the specimen that was to be identified. This method produced artificial groupings; the Aptera (arthropods without wings) contained such distantly related taxa as fleas, collembolans, arachnids, and certain crustaceans.

CRITERIA OF ZOOLOGICAL CLASSIFICATION

The procedure of classifying animals consists of combining "related" species into groups called taxa. A *taxon* is defined as *a monophyletic group of populations or taxa that can be recognized by their sharing of a definite set of characters; such a group must be sufficiently distinct to receive a name and to be ranked in a definite taxonomic category.* Although one demands that a taxon be recognizable by sharing a definite set of attributes, this set may be a set of one. Many taxa are recognizable by a single diagnostic character, such as the possession of a notochord in the chordates. It is assumed that all the members of a taxon are derived from the nearest common ancestor.

How does the taxonomist find the characters that group species into higher taxa? Similarity was the exclusive yardstick in the delimitation of taxa in pre-Darwinian days. Darwin established an entirely new theoretical foundation for biological classification: the genealogical principle of common descent. He emphasized correctly that the development of groups of species is a result of evolution and that the delimitation of taxa must be based on this recognition. Only groups of species related by common descent should be recognized as taxa. Difficulties inherent in the definition of common descent are discussed in Chapters 9 and 10.

Hennig (1950) rightly praised the "process of reciprocal illumination," in which a tentative classification produces a deeper understanding of the information content of the characters, possible homoplasies, and other information that permits the production of an improved revised classification. This, of course, corresponds to the hypothetico-deductive approach (Hull 1967). In the construction of a classification, the first step ordinarily consists in recognizing, by inspection, seemingly "natural" groups on the basis of similarity. The "naturalness" of these groups is subsequently tested by all available methods, including the presence of appropriate synapomorphies, the elimination of conflict with the fossil record, and the unmasking of homoplasies and other factors affecting judgment on relationship. The primacy of a first phenetic approach was recognized by Hennig (1966*a*) and has been stressed by Simpson (1961), Wagner (1980), and many other taxonomists.

CLASSIFICATION AND PHYLOGENY

In spite of Darwin's clear statement that classifications should be based on genealogy, his advice was difficult to follow, primarily because of a continuing uncertainty about how to recognize the descendants of a common ancestor. The first to implement Darwin's recommendation was Ernst Haeckel (1866), who claimed to have based his classifications on the phylogeny of the groups concerned. His opponents objected to this method by asking, How can we know the phylogeny? Is the truth not exactly the reverse of Haeckel's claim? Is not the phylogeny inferred from the findings made during the establishment of classifications?