Stephan Körner
University of Bristol, England

Classification Theory[1]

Körner, S.: **Classification theory.**
In: Intern. Classificat. 3(1976) No.1, p.3–6, 13 refs.

Some theory of classification underlies most practical activities, whether individual or social. It is involved in most attempts to understand the structure of reality on a metaphysical level and it is presupposed in the formulation of scientific laws. The logical and material principles of classification are covered as well as various general problems and criteria of classification. Differing classification principles in particular domains such as the natural sciences and medicine and in information are discussed as well as the place and role of classification in scientific method, its relation to and dependence on theory, scientific nomenclature and the philosophical issues through the heritage of Plato and Aristotle. (I. C.)

0. Introduction

In apprehendly the world, men constantly employ three methods of organization, which pervade all of their thinking: (1) the differentiation of experience into particular objects and their attributes–*e. g.*, when they distinguish between a tree and its size or its spatial relations to other objects; (2) the distinction between whole objects and their component parts–*e. g.*, when they contrast a tree with its component branches; and (3) the formation of and the distinction between different classes of objects–*e. g.*, when they form the class of all trees and the class of all stones and distinguish between them. Of these methods, the differentiation of objects and attributes is obviously presupposed by the other two. Though the whole-part and the class-member relationships are quite different, the work of developmental psychologists has indicated that children below the age of five cannot distinguish between them. This article, however, deals only with the third method.

Most practical activities, whether on an individual or social level, involve classification. The buying and selling of commodities (such as carloads of melons), for example, often concerns objects considered as members of a class (melons) rather than as concrete particulars. Classification is no less involved in any attempt at a theoretical understanding of the whole of reality or of some aspects of it. Ancient and recent metaphysicians, in their efforts to determine the structure of reality, have put forward classificatory schemes that allegedly reflect this structure. Formulation of scientific laws presupposes classifications, because to formulate a law of nature is to state relations between the members of different classes.

(1) We gratefully acknowledge the author's and publisher's kind permission to reprint this article from The Encyclopaedia Britannica, 15th Ed., Vol. 4 (Macropaedia)

1. The principles of classification

1.1 Logical principles

From the purely logical point of view, a classification of a domain of things does not depend on the nature of the criteria for class membership. It coincides with what, in the mathematical theory of sets, is called a "partition": a division of a set of objects into subsets is a partition if and only if

1. no two subsets have any element in common and
2. all of the subsets together contain all of the members of the partitioned set;

i. e., they are mutually exclusive and jointly exhaustive. A classification or partition may be refined by classifying or partitioning the subsets and their subsets until (if ever) a class of only one member is reached. If a set is manageably finite, its partition can proceed without employing any criteria for class membership by simply forming collections that satisfy the two conditions for a partition; *e. g.*, when the set {a, b, c, d} is subdivided into the subsets {a, b} and {c, d}. If a set is infinite or finite but unmanageably large, then its partition requires the use of criteria; *e. g.*, when the set of integers {1, 2, 3, . . .} is partitioned into the subsets of even and of odd integers. A criterion for class membership may be either a simple characteristic (*e. g.*, being an even integer) or a compound characteristic (*e. g.*, being divisible by 2 and by 3 or being divisible by 2 or by 3) so that possession of the characteristic is a necessary and sufficient condition for an object's membership in the class.

The mathematical theory of sets, however makes the unrealistic assumption that every set is exact or extensionally definite. It disregards the frequent occurrence of borderline cases; *i. e.*, of objects that can with eqal correctness be accepted or rejected as members of a class. Such borderline cases, common to two otherwise exclusive classes, are relevant in biological classification. A logical theory that allows for inexact classes has been developed for analyzing the relation between mathematical and perceptual propositions.

1.2 Material principles

Though governed by the same formal principles, classifications may differ widely in their classificatory criteria and in the principles determining their choice. It is usual to distinguish between natural and artificial, between essential and empirical, and between pragmatic and otherwise-justified classifications.

1. The distinction between natural and artificial classifications is hardly an absolute one: it is relative with respect not only to different cultures but also to different phases in the history of one culture; and this relativity applies even if a natural classification is defined by classes the members of which share the maximal number of

attributes. To a contemporary Westerner, for example, the classifications employed by the members of some primitive tribe – of days into auspicious and inauspicious, for example – may seem wholly artificial. Again, those of St. Thomas Aquinas, which contain a class of angels, may seem equally unnatural to him.

2. The distinction between essential and empirical classification is based on the assumption that the former rests on a priori ideas as to what is important, whereas the latter rests on observation alone. Yet no scientific classification is independent of theoretical assumptions as opposed to uninterpreted observations, if, indeed, there are such things. To regard, for example, zoological classifications that are not genetic as wholly nonempirical and those that are genetic as wholly empirical is to mistake a change of theory for a discovery of an error.

3. Pragmatic classifications in the sense of philosophical Pragmatism (*q. v.*) must be distinguished from pragmatic classification meant to be merely provisional, heuristic (aiding discovery), auxiliary, or made independently of scientific theorizing.

2. The domains of classification

2.1. General problems

In every attempt at classifying a domain of objects, the extent to which the choice of classificatory principles depends upon the nature of the objects must be considered. More specifically, the choice of the principles may depend, as in acoustics, on the extent to which the objects of the domain are given in perception; as in paleontology, on the extent to which they are subject to change or development; as in petrology, on the extent to which their differences are differences in degree rather than in kind; or, as in fluid dynamics, on the extent to which their differences are differences in quantity rather than quality.

Classification of perceptual and nonperceptual objects. In forming classes of perceptual objects – *e. g.*, the class of green things, of elephants, or of motorcars – the perceptual resemblances and dissimilarities between their members play an important role. Whatever definition of such a resemblance class may be adopted, it must always satisfy the following requirements: (1) the qualifications and disqualifications for membership must include a method for exhibiting standard members and nonmembers of the class, such that (2) an object qualifies for membership only if it is sufficiently similar to the standard members and sufficiently dissimilar to the standard nonmembers. Although the latitude allowed by these conditions can be restricted by various means, it cannot be wholly eliminated; thus, resemblance classes are inexact; *i. e.*, they admit of borderline cases. Their existence, far from impairing the classificatory scheme, may be a logical consequence or a postulate of a scientific theory employing the scheme. After all, if, say, the development of living organisms implies gradual change, their classification would be unrealistic if it did not allow for cases on the borderline between species. On the other hand, many scientific theories, such as those of theoretical physics, do not refer directly to perceptual phenomena but do so indirectly by relating a perceptual domain to a domain of abstract or ideal objects; *e. g.*, of Newtonian particles or Maxwellian fields. Such objects are described and classified by means of nonperceptual, structural properties and relations and expressed in the language of a mathematical theory – especially algebra, the theory of functions, and topology.

Classification by morphological and genetic criteria. A domain of objects that are unchanging or the history of which is regarded as negligible is classified only in terms of form or structure; *i. e.*, morphologically. Thus, Christian biologists of the 18th century would hold that animal species are constant, having been created constant; and some contemparary anthropologists would hold that the history of a primitive tribe does not affect its basic social structure. If, on the other hand, the domain of classification consists of developing populations of plants, animals, or stars, then the criteria of classification are likely to be genetic; *i. e.*, to refer to what are regarded as crucial development stages. Whereas a morphological classification need not be genetic, any genetic classification must be to some extent morphological.

Classification by differences of kind and of degree. Sometimes the objects of a classificatory domain differ from each other not so much in their characteristics as in the degree to which they possess them. Thus, minerals may be classified according to their increasing hardness and commodities by the increasing preference shown for them by the buying public. A classification of this type is or is based on a so-called partial ordering. More precisely, a domain of objects is partially ordered by a relation – say, $\leqslant$ ("smaller than or equal to") – if and only if, for any objects x, y, z of the domain, (1) $x \leqslant x$, (2) $x \leqslant y$ and $y \leqslant x$ implies that $x = y$, and (3) $x \leqslant y$ and $y \leqslant z$ implies that $x \leqslant z$. In technical language, the relation must be reflexive (holding between an object and itself), antisymmetrical (the applicability of both the relation and its converse implying the identiy of the terms), and transitive (as in descendance, which implies that the descendant of the descendant of some forebear is *ipso facto* a descendant of this forebear). That a partial ordering involves a classification is especially clear where a larger number of objects are equal with respect to the ordering relation; *e. g.*, belong to a fairly large class of minerals of equal hardness or to a class of commodities none of which the buying public prefers to any other. A partial ordering is total if and only if the ordering relation is dichotomous; *i. e.*, if, for all x and y, $x \leqslant y$ or $y \leqslant z$.

Classification by differences of quantity and of quality. Quantitative measurement, as opposed to mere ordering, establishes equalities and inequalities of order or rank not only between different single members of a domain but also between different pairs of them. It allows a scientist, for example, not only to order objects by their temperature but also to order the differences in temperature between any two of them. Just as the ordering of a domain establishes classes of objects equal in rank, so also its measurement establishes classes of objects equal in quantity. Though many sciences have tended to develop from mere qualitative comparison and classification toward ordering and measurement, the adage that science is measurement is an exaggeration.

2.2 Classification in particular domains

Principles of classification depend to some degree upon the domain involved.

Classification in the natural sciences. The greater the role played by purely quantitative methods, the smaller that played by merely qualitative classifications. Hence, comparatively less attention is given to classification in the physical than in the biological sciences. In the more descriptive parts of physics, however, classification is still of utmost importance. In astronomy, for example, difficulty is experienced not only in determining characteristic features (as in the case of the galaxies) but also in making sure that their observability is not lost as increasingly distant objects are studied. In moving from physics through chemistry to biology, the role of classification becomes more dominant; and, in biology, taxonomy, or the ordering of organisms into species, genera, families, and so on, constitutes a central part of the theory.

Classification in the social sciences. Classification in the social sciences was and still is to some extent concerned with so-called ideal types, such as the "typical bureaucrat", limiting concepts, which, though not exemplified in reality, serve nevertheless to explain the social behaviour of real people by concentrating on and even exaggerating certain features of people while ignoring others. Though the predominance of ideal types in the social sciences may simply mark an early stage in their development, whether they are now dispensable is controversial. From the logical point of view, a classification into ideal types is a classification of real people only insofar as real people can be ordered by the degree to which they approximate the type. And, more generally, a classification into ideal phenomena requires for its application and ordering of real phenomena.

Classification in the applied sciences and medicine. Although the distinction between pure and applied sciences – say, between zoology and animal husbandry – is not sharp, the latter are more concerned with practical than with theoretical ends. Thus, a rough classification of a domain – say, of different building materials or of different strains of a virus – may be preferable to a finer classification if their practical utility – say, for the building of bridges or the curing of diseases – is the same; or if, relative to their respective utility, the cost of the rough classification is very much lower than that of the fine.

Classification of information. When the purpose of classification is simply to make information available, the predominance of purely practical ends over theoretical is even more marked. Thus, the purpose of library classification is not so much to exhibit the fundamental relations among the things classified as it is to exhibit relations that are helpful in locating the information being sought. It would seem futile to argue, for example, whether "coal mining" should be a subdivision of "mining" or of "coal" (there are actual systems that do it each way). Similar problems arise for the classification schemas underlying encyclopaedias such as the present work, which aims at treating every existing subject. When information is stored by computers, the usual principles of classification are modified by those governing the technology of computers.

3. The place and role of classification in scientific method

3.1 Classification: its relations to and dependence on theory

Though purely classificatory sciences are sometimes contrasted with explanatory sciences, it must be emphasized that the formulation of scientific laws presupposes classification. This is true not only of universal laws of nature but also of probabilistic laws. As *R. B. Braithwaite*, a British philosopher of science, has emphasized, every deterministic scientific generalization may be (at least partly) analyzed as a concomitance generalization to the effect that everything that is A is B – provided that A and B are sufficiently complex properties – and, clearly, the principles for setting up the classes A and B in the first place must serve as a basis for the generalization.

Probabilistic or statistical laws of nature also presuppose classification, because any such law has the form of a statement that a certain proportion of things belonging to class A belongs to class B or that there is a certain probability that a thing that belongs to A also belongs to B. Universal laws that can be formulated within one classificatory scheme may not be amenable for formulation within another. And the same holds for statistical laws. Here the proper choice of the related classes is important: the mortality, for example, of people of ages 40 to 50 suffering from a certain disease is of interest but not that of people so aged whose Christian name consists of two syllables.

While every theory presupposes a classificatory scheme, this scheme is in turn influenced by the content of the theory. This influence is perhaps most obvious in biology, in which the transition from the pre-evolutionary to the evolutionary point of view has influenced taxonomy in several ways. First, the hypothesis that species are not fixed units but are entities that change and grade into each other has made it necessary to regard the extension of species as variable and as necessitating borderline cases. Second, the hypothesis that one species may descend from another as a result of organic reproduction has made it necessary to base the classification into species on the notion of a population of animals exhibiting a frequency distribution of certain characters.

The tendency to base classifications on frequency and probability distributions of variable characters within populations (or ensembles) rather than on homogeneous classes has been manifest also in theoretical physics ever since quantum mechanics was developed as an irreducibly statistical theory. Whereas before the advent of quantum mechanics statistical hypotheses were regarded as compatible with and, at least in principle, reducible to universal laws, the opposite point of view is now dominant. Thus, the physical and biological sciences reinforce each other in implying that the theoretically most basic scientific classifications depend on statistical distributions of variable characteristics rather than on constant criteria.

A similar shift toward classification in terms of statistical distributions can also be noticed in the social sciences, in which, as *Paul Lazarsfeld*, a communications sociologist, has emphasized, the investigator will frequently have to develop his own classificatory scheme rather than to take one over from a developed, explicit theory.

The place of the theory is taken by a provisional model or scheme of the whole situation in which the inquiry has taken place. Use of such a model suggests that a classificatory scheme is required that, when modified as a result of the inquiry, will in turn suggest modifications of the model. The distinction between classifications based on explicit theories and classifications suggested by structural models is, of course, not sharp. And, again, the latter kind of classification cannot be sharply distinguished from those based on a more or less implicit sense of proportion or reasonableness.

3.2 Classification and scientific nomenclature

The more complex a classificatory scheme, the more difficult is its application and the more important the choice of a suitable terminology and nomenclature. These problems are particularly pressing in biology, in which, as the leading evolutionist *G. G. Simpson* points out, the existence of millions of species is acknowledged, each of which must be named – quite apart from more general and less general classes. The subjective and arbitrary element in the choice of a system of nomenclature is recognized by the organization of international congresses to arrive at agreements on conventional names. Objectively, the taxonomically most important features must also be emphasized in the system of nomenclature. Thus, according to Darwin, those characters that, in the course of evolution, have suffered the least modification are taxonomically most important and should be given a central place in any system of nomenclature. The history of the transition from Linnaean to Darwinian and post-Darwinian theory illustrates the dependence of nomenclature on taxonomy and of taxonomy on theory. At the same time it also shows how an established nomenclature tends to preserve established taxonomical principles and thus indirectly to perpetuate the theory on which they are based.

3.3 Philosophical issues regarding classification

From the rise of philosophical reflection, some classifications have been viewed as adequate to reality and others as erroneous. Plato's theory of Forms, the earliest metaphysical theory of classification, is still the paradigm of all typological classifications. The Platonic Forms are unchanging ideal objects – in particular, mathematical objects – by reference to which the fluctuating objects of sense experience are classified and ordered. Perceptual objects and the relations between them are not instances of Forms or of relations between Forms but only participate in or approximate them. In asserting that one apple and one apple make two apples, one asserts that perishable perceptual objects approximate eternal mathematical units and that a physical operation involving perishable objects approximates a mathematical relationship.

Aristotle rejects the Platonic Forms and the relationship of participation in favour of the relationship between attributes and their instances. The Aristotelian theory of classification and of definition by classification has both an uncontroversial logical aspect and a controversial metaphysical aspect. A definition formulated by classification of kinds of things consists, according to Aristotle, in indicating a simple or compound attribute that the defined kind shares with other kinds and by indicating another such attribute that it does not share with the other kinds. A definition by classification is also called a definition by *genus proximum* ("next-higher genus") and *differentia specifica* ("specific difference") – a nomenclature especially apt if one assumes, with Aristotle, that the correct choice of *genus* and *differentia* is not dependent on convention or convenience but on the nature of reality. It is held by some theorists that there is one and only one adequate classificatory hierarchy such that each kind of thing, unless it is a lowest kind (*infima species*), is divided into two or more lower kinds (*species*) and that each kind of thing, unless it is a highest kind (*summum genus*), falls under one higher kind – a view that is sometimes called essentialism, because it bases the classification of things on their alleged essences. Some form of the essentialist doctrine that there must be one essentially natural system of classification is held by most metaphysicians, who thus assume that whatever exists falls into one or more natural kinds (*e.g.*, minds, bodies, or minds and bodies).

The essentialist, doctrine is clearly rejected by *W. S. Jevons*, one of the founders of modern symbolic logic and philosophy of science. He devotes a whole chapter to classification, the value of which he regards as "co-extensive with the value of science and general reasoning." His careful investigation into the employment of classification in the different branches of science, which is as modern today as in 1874, convinces him that there is no unique, essential, natural, or a priori system of classification that is alone adequate to the nature of reality.

This conclusion is compatible with the possibility and, indeed, the historical fact that at some period of time a certain classificatory scheme or part of one that is actually employed may appear to its users to be more adequate than any alternatives and thus to be incorrigible.

Annotated References

1 Kuratowski, K., Mostowski, A.: Teoria mnogości. 2nd ed. 1966. (Engl. Transl.: Set Theory, 1967 – one of many excellent books on set theory)
2 Körner, S.: Experience and Theory. 1966. (Contains elaborate discussions of inexact classes and empirical continuity)
3 Strauss, C. L.: La Pensée sauvage. 1962. (Contains an anthropological theory of the nature and function of classification)
4 Simpson, G. G., Pittendrigh, C. S., Tiffany, L. H.: Life. 1957. (An excellent survey)
5 Hempel, C. G.: Aspects of scientific explanation and other essays in the philosophy of science. 1965. (Contains one of the best discussions of typological classification)
6 Hubble, E. P.: The realm of the nebulae. 1936; repr. 1958. (A standard book)
7 Mills, J.: A modern outline of library classification. 1960. (An informative survey)
8 Braithwaite, R. B.: Scientific explanation. 1953. (A clear discussion of most problems in the philosophy of science)
9 Dirac, P. A. M.: The principles of quantum mechanics. 1930. (A fundamental work)
10 Lazarsfeld, P. F., Rosenberg, M. (Eds.): The Language of Social Research: A Reader in the Methodology of Social Research. 1955. (Contains informative methodological papers)
11 Cain, A. J. (Ed.): Function and Taxonomic Importance. 1959. (A modern discussion of classification theory)
12 Jevons, W. S.: The Principles of Science (esp. p. 305–315). 1874. (A classic)
13 Woodger, J. H.: Biology and Language. 1952. (Mainly concerned with logical questions)