# SCIENTIFIC AMERICAN

Numerical Taxonomy Author(s): Robert R. Sokal Source: *Scientific American*, Vol. 215, No. 6 (December 1966), pp. 106-117 Published by: Scientific American, a division of Nature America, Inc. Stable URL: https://www.jstor.org/stable/10.2307/24931358

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms



Scientific American, a division of Nature America, Inc. is collaborating with JSTOR to digitize, preserve and extend access to Scientific American

## Numerical Taxonomy

The computer has made it possible to consider large numbers of characteristics in classifying many phenomena, notably living organisms, fossil organisms and even imaginary organisms

#### by Robert R. Sokal

Classification is one of the fundamental concerns of science. Facts and objects must be arranged in an orderly fashion before their unifying principles can be discovered and used as the basis for prediction. Many phenomena occur in such variety and profusion that unless some system is created among them they would be unlikely to provide any useful information. Chemical compounds (particularly organic ones), groups of stars and the two million or so species of living organisms that inhabit the earth are examples of such phenomena.

The development of high-speed electronic computers has had a profound impact on the methods of classification in many scientific fields. The rapidity of the computer's operation has made it possible for the first time to consider large numbers of characteristics in classifying many phenomena. The writing of computer programs for such work has led to a renewed interest in the principles of classification, reviving such old questions as: What makes one classification better than another? What is a "natural" classification? What is similarity, and can it be quantified? The inquiry has progressed furthest in the field of taxonomy, or biological classification. The methods of numerical taxonomy (as this new field has come to be called), the conceptual revolution it has wrought, the nature of the controversy surrounding it, some future prospects for the field and its relevance to problems of classification in other sciences will be discussed in this article.

Many of the new procedures of numerical taxonomy and their theoretical justification have been the subject of intense disagreement between numerical taxonomists and supporters of traditional taxonomic practices and principles. Controversy, of course, is nothing new in science. Time and again the introduction of a new concept or the development of a new technique has aroused the passions of scientists representing conflicting points of view. Although debate about numerical taxonomy has not been as acrimonious as some debates in the history of science, it has certainly been spirited and continues undiminished. At recent biological conferences the symposiums on numerical taxonomy have been unusually well attended, often by people only remotely interested in the field who have heard that "a



IMAGINARY ANIMALS, called Caminalcules after their creator, Joseph H. Camin of the University of Kansas, are used in experiments on the principles and practices of tax-

106

good fight" was about to take place in that session. What is all the shooting about?

In the early days of modern science, and for special purposes even today, classifications were based on a single property or characteristic, the choice of which might be quite arbitrary. Metals are divided into conductors and nonconductors, other substances into those that are soluble in water and those that are not; organisms are divided into unicellular ones and multicellular ones. Some of these classifications are arbitrary in the sense that there is a continuum of properties-as in the case of solubility, for which the line between soluble substances and insoluble ones is not distinct. In contrast one can almost always say whether an organism is unicellular or multicellular, so that with properties such as these the decisions can be quite clear-cut. Classifications based on one or only a few characters are generally called "monothetic," which means that all the objects allocated to one class must share the

character or characters under consideration. Thus the members of the class of "soluble substances" must in fact be soluble.

Classifications based on many characters, on the other hand, are called 'polythetic." They do not require any one character or property to be universal for a class. Thus there are birds that lack wings, vertebrates that lack red blood and mammals that do not bear their young. In such cases a given "taxon," or class, is established because it contains a substantial portion of the characters employed in the classification. Assignment to the taxon is not on the basis of a single property but on the aggregate of properties, and any pair of members of the class will not necessarily share every character.

It is obviously much more complicated to establish classifications based on many characters than it is to establish classifications based on only one character. The human mind finds it difficult to tabulate and process large numbers of characters without favoring one aspect or another. The comparative subjectivity of traditional approaches and the inability of taxonomists to communicate to one another the nature of their procedures have contributed to making taxonomy more of an art than a science.

The arrival of the computer has reversed this trend, and a new field with many possibilities for objective and explicit classification has opened up. Computer techniques have indeed been a principal force behind the gradual adoption of an operational approach in taxonomy; in order to use such techniques, classificatory procedures must be outlined in such a form that any scientist or a properly programmed computer can carry out the indicated operations and, given the same input data, arrive at the same results. This would preclude the often arbitrary decisions of conventional taxonomists, epitomized by the statement that "a species is whatever a competent taxonomist decides to call a species."

Before proceeding further I should remove a possible source of confusion.



onomy, or biological classification. The 29 "recent" species of the organisms, depicted on these two pages, were generated by Camin

according to rules known so far only to him. The drawings are based on Camin's originals, with slight modifications in perspective.



TAXONOMIC RELATIONSHIPS can be viewed from three distinct aspects. "Phenetically" (based on overall similarity among the objects to be classified) organism B is more closely related to organism A than it is to organism C, even though C evolved much later than A as a branch of stem B. "Cladistically" (based on common lines of descent) organisms B and C are closer to each other than either is to A, since they have an ancestor (B')

before either has a common ancestor (A') with A. "Chronistically" (based on time) A, B and C are closer to one another than any of them is to B', since they occupy same time horizon.



SIMILARITY CAN BE REPRESENTED as the distance between the objects to be classified (called operational taxonomic units, or OTU's for short) in a multidimensional space. In this example the similarity between all possible pairs taken from four objects is estimated on the basis of three characters, which are represented by the three coordinates axes X, Y and Z. The OTU's (*black balls*) are then plotted into this three-dimensional space according to their state, or value, for the three characters. Similar objects are plotted much closer to one another than dissimilar ones. In any real case there will, of course, be more than three characters and a multidimensional space—called a "hyperspace"—would be necessary.

This is the difference between the terms "classification" and "identification." When a set of unordered objects has been grouped on the basis of like properties, biologists call this "classification." Once a classification has been established the allocation of additional unidentified objects to the correct class is generally known as "identification." Thus a person using a key to the known wild flowers of Yellowstone National Park "identifies" a given specimen as a goldenrod. Some mathematicians and philosophers would also call this second process classification, but I shall strictly distinguish between the two. Here I am principally concerned with classification in the biologist's sense.

The purpose of taxonomy is to group the objects to be classified into "natural" taxa. Naturalness has been variously defined, but underlying the several definitions is the common idea that members of a natural taxon are mutually more highly related to one another than they are to nonmembers. This leads us to try to define what we mean by "taxonomic relationship." Conventional taxonomists wish to equate taxonomic relationships with evolutionary relationships, but numerical taxonomists have pointed out that taxonomic relations are actually of three kinds. "Phenetic" relationships are those based on overall similarity among the objects to be classified. "Cladistic" relationships are based on common lines of descent. Although close cladistic relationship generally implies close phenetic similarity, it is not always the case. Differences in evolutionary rates may give rise to lineages that diverged long ago but appear more similar than a subsequently diverged pair of stems, one or both of which has undergone rapid evolution [see top illustration at left]. The third kind of taxonomic relationship is the "chronistic," or temporal, relation among various evolutionary branches. Cladistic relationships for most organisms are known scantily, if at all, and are generally inferred from phenetic evidence. The "phylogenetic" classifications of conventional taxonomy are usually based on an undefined mixture of phenetic and cladistic relationships, and often merely represent an overall similarity among the classified organisms disguised in evolutionary terminology.

In view of these considerations numerical taxonomists propose to base classifications entirely on resemblance, defining natural classifications as those

yielding taxa whose members are in some sense more similar to one another than they are to members of other taxa. It follows from this concept of naturalness, which is based on the ideas of J. S. L. Gilmour, a botanist at the University of Cambridge, that a natural taxon will be most predictive. If a classification is based on many correlated characters, predictions about the states of other characters in various groupings of the classification should be more successful than if the taxonomy were based on few characters. Furthermore, it is likely that a classification based on a great variety of characters will be of general utility to biology as a whole, whereas a classification resting on only a few characters is less likely to be generally useful, except for the special purposes relevant to the chosen characters. Thus a classification of animals into "swamp dwellers" and "animals not living in swamps" may be very useful for a study of the ecology of swamps but not for general zoology.

Overall phenetic similarity is based on all available characters without any differential weighting of some characters over others. A substantial part of the controversy about numerical taxonomy has centered on this point. Conventional taxonomists usually employ only a few characters in classification and weight these in terms of their presumed evolutionary importance. Numerical taxonomists contend that evolutionary importance is undefinable and generally unknown and that no consistent scheme for weighting characters before undertaking a classification has yet been proposed. To weight characters on the basis of their ability to distinguish groups in a classification, as is frequently advocated, is a logical fallacy. Since the purpose of employing the characters is to establish a classification, one cannot first assume what these classes are and then use them to measure the diagnostic weight of a character.

The nature of similarity is, of course, a fundamental problem of taxonomy, whatever one's theoretical approach. This ancient philosophical problem has recently become acute in a variety of fields because of the introduction of automata for classification and identification. What is the meaning of the statement "A is similar to B"? Only when qualified to the effect that "A is similar to B in such and such a respect" has this statement any meaning. It is one of the underlying assumptions



PHENOGRAM is a convenient two-dimensional representation of the results of a numerical classification, in this case the results of classifying the 29 recent species of Caminalcules depicted on pages 106 and 107. The various species are indicated by the numbers at the tips of the branches. Phenograms tend to distort the original multidimensional relationships.



DISSIMILARITY (PHENETIC DISTANCE)





of numerical taxonomy that quantification of degrees of similarity is possible. The establishment of comparative similarities—for example "A is more similar to B than it is to C"—is fundamental to any attempt at clustering objects into homogeneous groups.

Similarity can be established only on the basis of homologous, or corresponding, characters. Hence it is not possible to compare the forelimbs of vertebrates without prior agreement on what to call a forelimb in each of the vertebrates to be compared, and on the correspondences between constituent parts of the appendages. Homology, as interpreted by numerical taxonomists, is the existing overall similarity in structure rather than similarity due to common ancestry, although this may often be the underlying cause. To describe such essential similarity one needs to base it on numerous "unit characters" of the structures to be compared. Numerical taxonomists regard unit characters as those that cannot be subdivided into logically or empirically independent characters. This is a complex subject, however, since the same set of biological characters can be described in innumerable slightly varying ways. One would not wish to use all these descriptions, yet how can one avoid redundancy by choosing the best ones?

Another problem is how many characters to choose for describing phenetic similarities. Is there an asymptotic similarity among organisms that is approached as more and more characters are measured, or will each additional set of characters contribute a new dimension to similarity, making the taxonomic structure of a group inherently unstable? All the evidence on this complicated question is not yet in. It might be assumed that if one knew the genetic fine structure of organisms, one could then develop an overall measure of similarity among organisms based on similarity of genetic structure.

SIMILARITY MATRIXES have been shaded to show the degree of similarity between pairs of 27 OTU's (in this case individuals from seven species of nematode worms). The darker the squares, the greater the similarity. The matrix at top has the OTU's arranged according to an arbitrary sequence of code numbers. The matrix at bottom has been rearranged to yield clusters of similar OTU's. The dark triangles along the diagonal indicate species; larger, less dark triangles represent genera. OTU 13 is not closely related to any of the other OTU's (see illustration on opposite page).

110

Yet even this would present complications, since the genetic code as it is now understood is in the nature of a program, certain portions of which come into play at different times during the development of an organism. Similarity in the programs might not reflect similarity in the products, and it is by no means certain whether genes or their effects should form the basis of a classification.

Moreover, since we do not as yet have measures of similarity between different genetic codes (except for certain limited instances), we are forced to resort to the morphological and physiological characters employed in conventional taxonomy. Recently we have found that although different types of characters in a taxonomic study may be correlated, this correlation is not sufficiently strong for a classification based on one set of characters (for example external characters) to agree fully with a classification based on a second set (for example internal characters). Thus a taxonomy of males may differ somewhat from one of females, and a classification of skeletal parts may not agree entirely with one based on soft parts. This is a necessary consequence of phenetic classification, and in order to obtain valid measures of overall similarity one has to use as many and as varied sets of characters as possible.

If classifications are to be established on overall similarity, numerical taxonomy is required to put the procedures on an operational and quantitative basis. Some of the procedures of numerical taxonomy were developed as early as the beginning of this century, but before the introduction of digital computers they never caught on, presumably because of the insuperable computational difficulties. The philosophical origins of the present development in taxonomy derive from the work of Michel Adanson, an 18th-century French botanist, who first rejected a priori assumptions on the importance of different characters and proposed basing natural taxa on his essentially phenetic concept of "affinity."

The recent development of numerical taxonomy starts with the almost simultaneous publication in 1957 of papers advocating this method by Peter H. A. Sneath, a British microbiologist, and by Charles D. Michener and myself, both entomologists at the University of Kansas. Two further independent studies by workers at the University of Ox-



NEMATODE PHENOGRAM is based on the similarity matrix at bottom of opposite page. The brackets and lettering correspond to similar brackets and lettering in the similarity matrix. Broken vertical lines are "cutoff lines" for recognizing species and genera, which are indicated by their full names at right. Code numbers for OTU's are at tips of branches. Rearrangement of first two species has no effect on the taxonomic relationship illustrated.

ford and at the New York Botanical Garden followed in 1958 and 1960 respectively. Since that time the literature and the number of workers in the field have grown rapidly. At last count there were at least 200 published papers on numerical taxonomy, with more than 60 papers applying numerical taxonomy to diverse groups of organisms.

How does one produce a classification by numerical taxonomy? The objects to be classified are called "operational taxonomic units," or OTU's for short. They may be individuals as such, individuals representing species or higher-ranking taxa such as genera or families of plants or animals, or statistical abstractions of the higher-ranking taxonomic groups.

Classifications by numerical taxonomy are based on many numerically



DIFFERENT CLUSTERING TECHNIQUES yield different classifications of the same taxonomic units. The numbered black dots represent eight species of mosquitoes described by pupal characteristics only (*drawings at top*). Species 30 belongs to the genus *Mansonia*; the others are species of *Anopheles*. For convenience of representation, distances among OTU's are shown in two-dimensional space only. Linkages between OTU's and clusters and between pairs of clusters are shown by solid-colored lines, previously established clusters by light-colored shading of the area occupied by the clustered OTU's. As clustering proceeds step by step the criteria for joining become less stringent; in other words, the distances between prospective joiners and established clusters increase. The single-linkage method (*left*) starts with the shortest distance between any pair of OTU's and takes up the other distances in order of magnitude. In average linkage (*center*) an OTU will join a cluster if the average distance between it and the "center of gravity" of the cluster is less than for any other such distance in the study. In complete linkage (*right*) joining takes place only when the relationships between a candidate for joining and established members of the clusters are all at the minimum criterion for a given clustering cycle. Although the initial clustering step is the same for all three methods and the final cluster, including all the OTU's, must necessarily also be the same, the intermediate clustering steps are obviously quite different at roughly equivalent stages. The results depicted here will not necessarily agree with comparable studies that include additional closely related OTU's.

recorded characters. These may be measurements that are appropriately represented numerically, or they may be coded in such a way that the differences between them are proportional to their dissimilarity. For example, a character called "hairiness of leaf" might be coded as follows: hairless, 0; sparsely haired, 1; regularly haired, 2; densely haired, 3. By this coding system we imply that the dissimilarity between densely haired and hairless is approximately three times the dissimilarity between sparsely haired and hairless. In some fields, such as microbiology, characters are almost always expressed by only two states corresponding to the presence (1) or the absence (0) of a given character, for example an enzyme.

All the characters and the taxonomic units to be classified are arranged in a data matrix, and the similarities between all possible pairs of OTU's are then computed based on all the characters. We shall not concern ourselves here with the variety of mathematical coefficients that have been devised to represent similarity between objects. One way of representing similarity (actually dissimilarity) is the distance between OTU's in a multidimensional space. Suppose the similarity between all possible pairs taken from four objects is to be estimated on the basis of three characters. We can visualize these characters as representing three coordinate axes [see bottom illustration on page 108]. Each OTU is then plotted into this three-dimensional space according to its state, or value, for the three characters. Those objects that are very similar will be plotted close to each other; dissimilar ones will be considerably farther apart. The computation of such straight-line distances is quite simple. In any real case there will, of course, be more than three characters and a multidimensional space would be necessary. Although it is not possible to represent such a "hyperspace" pictorially, the computation of distances within it is still quite simple. Thus we can view the objects to be classified as clusters of points in multidimensional space.

The similarities between pairs of OTU's are evaluated by a computer and printed out in a "similarity matrix," which shows the similarity value of each OTU with respect to every other one. Rather than give such a numerical table here, I have illustrated it graphically on page 110, indicating the magnitude of the similarity coefficient by depth of shading. Unless the OTU's to



GENETIC CONTINUITY was accomplished in the generation of the Caminalcules by tracing the drawing of the primitive species (*bottom*) from sheet to sheet, making possible the preservation of all characters except for the desired morphological modifications (*color*).

be classified have been ordered previously, the pattern of shading in the similarity matrix is likely to be complex. We can attempt, however, to alter the arrangement of the OTU's in such a way that the dark-shaded areas (highsimilarity values) will condense in triangular groups along the diagonal of the table. This procedure will yield a rough classification of these OTU's into groups.

For more precise classifications a variety of numerical clustering procedures have been developed, and these procedures are routinely carried out on the computer after the similarity matrix has been calculated. There is no generally accepted clustering method. Different methods will yield different results, depending on the underlying "similarity structure" of the objects to be clustered [see illustration on opposite page]. Attempts are being made currently to define an "optimal" classification mathematically so that the results of a numerical classification can be evaluated by this criterion.

The results of a numerical classification are usually represented by means of a "phenogram." These treelike diagrams indicate the similarity between OTU's or stems bearing more than one OTU along one axis. Because phenograms collapse multidimensional relationships into two dimensions, there is appreciable distortion of the original relationships as shown in the similarity matrix. Estimates of the degree of distortion in a given phenogram are made routinely in numerical taxonomic studies as a precaution. Representing phenetic relationships by three-dimensional models of OTU's avoids some of the distortions encountered in phenograms. Since such models cannot be circulated widely, the possibility of publishing computer-produced "stereograms"—two-dimensional projections of three-dimensional models—is currently being investigated.

Describing the similarities among organisms is only one aim of taxonomy. Another is to trace the evolutionary lineages that gave rise to the diversity of organic life that exists today. To reconstruct the taxonomic relationships and evolutionary trends among a group of organisms one would need to describe their phenetic relationships through all points in time. One would also have to describe the group's "cladistics," the branching sequences in the evolutionary trees. Finally, one must furnish a correct time scale to the evolutionary reconstruction. At the moment there is no known way-short of a multidimensional reconstruction, which is impossible of practical achievementto incorporate these elements into a unified system without large distortion of the phenetic relationships.

Some substantial recent advances in

© 1966 SCIENTIFIC AMERICAN, INC



CLADOGRAMS, diagrams that delineate the branching sequences in an evolutionary tree, are shown here for a group of fossil horses. A computer program developed by Camin and the author constructs cladograms with the fewest number of evolutionary steps. The cladogram at top shows an early stage in the procedure; the one at bottom shows the most parsimonious solution, in which the 35 steps at top have been reduced to 31 steps. The cladogram at bottom corresponds to evolutionary branching sequence generally accepted by paleontologists. The evolutionary steps for various numbered skeletal and dental characters are marked on the branches. In the bidirectional evolution of a character one direction is shown by lines across branches, the other by X marks.

techniques for reconstructing cladistic sequences grew out of an experiment on the principles and practices of taxonomy carried out by a group of graduate students and faculty members at the University of Kansas. The study was based on a group of imaginary animals generated by Joseph H. Camin according to rules known so far only to him but believed to be consistent with what is generally known of evolutionary principles. Genetic continuity was accomplished by tracing the drawings of the animals from sheet to sheet, permitting the preservation of all characters except for the modifications that were desired. All 29 "recent" species of these "organisms," irreverently named Caminalcules by the graduate students, are shown on pages 106 and 107. Detailed studies of the assemblage of hypothetical animals by orthodox phylogenetic methodology by various team members resulted in differing, but internally consistent, cladistic schemes, the choice among which was not easily apparent. Comparison by Camin of these various schemes with the "truth" led him to the observation that those trees that most closely resembled the true cladistic sequence invariably required for their construction the fewest number of postulated evolutionary steps for the characters studied.

Our experiments were based on three working assumptions: first, that character states could be numerically coded according to their presumed evolutionary trends; second, that evolution is irreversible, so that when a character evolves to state 2, it will not revert to state 1; third, that nature is fundamentally parsimonious, so that the diversity in character states within a given group was achieved at or close to the minimum number of evolutionary steps. From these assumptions Camin and I developed a computational technique that constructs the most parsimonious cladistic tree, or cladogram, from an original data matrix. A computer program carries out these computations. The cladograms on the opposite page illustrate the type of change that is routinely carried out by the computer program. The cladograms estimate the branching sequences that occurred in the evolutionary history of a group of fossil horses. These methods have also given apparently meaningful results in studies of bees, vipers, certain plants, fossil protozoa and the structural rearrangements of chromosomes in blackflies and drosophila.

The computer program also evalu-



For Christmas and all of 1967, you couldn't choose a more perfect gift for your friends and associates than a subscription to SCIENTIFIC AMERICAN. Whether they are technical executives in industry, engineers, scientists, students or interested laymen, they will be enjoying their issues and thanking you for them long after most people have forgotten other gifts.

Take advantage of the low gift rates by sending your gift list today. If you indicate, each of your friends will receive an attractive Christmas card to inform him of your thoughtful gift.

P.S. If you are not now a subscriber, this is a timely opportunity for you to take advantage of this Christmas subscription offer.





AUTOMATIC SCANNING of organisms for the purpose of establishing phenetic classifications was proved to be feasible in a recent experiment performed by F. James Rohlf and the author in which they successively placed 25 punch cards, each perforated with 25 randomly chosen holes, over drawings of a group of Caminalcules. One of the punch cards is shown at left superposed over a drawing



of a Caminalcule belonging to species 1. In making the composite at right each of the 625 holes was scored 1 when a black line appeared through it and 0 when no black showed. (Actually fewer than 625 holes are visible here, since many holes on different cards coincided.) The illustrations of the organisms were compared on the basis of matching scores for corresponding masks and holes.

ates the compatibility of each characteristic with all the other characters and weights it in terms of this criterion. It points out inconsistencies and has repeatedly discovered errors in coding, transcription or interpretation of the data.

major impetus for the development A and application of numerical taxonomy is the current introduction of automatic sensing and data-recording devices. The development of such instruments has proceeded very rapidly in recent years. Most prominent among the devices likely to be useful in taxonomy are optical scanners, which digitize drawings, photographs, microscope preparations and results of biochemical analysis. The veritable flood of information that will flow from these automatic sensors will require computerbased processing and classification, since the human mind is not able to digest these data by traditional means.

Recently F. James Rohlf and I have shown that data of this kind, collected in a quite unsophisticated manner, can be used to form adequate phenetic classifications. We employed the straightforward approach of recording agreement in visible structures over randomly selected minute areas of the images of pairs of organisms. Such a procedure would be feasible by means of optical scanners. Random masks, made from 25 punch cards each perforated with 25 randomly chosen holes, were placed over black-and-white drawings of two groups of "organisms." One group consisted of the 29 recent species of the Caminalcules; the second comprised published illustrations of the pupae of 32 species of mosquitoes. Each illustration was overlaid with all the masks, and each of the 625 holes was scored 1 when a black line appeared

through it and 0 when no black showed [see illustration above]. Illustrations were compared on the basis of matching scores for corresponding masks and holes. A numerical classification of the images was surprisingly similar to studies by conventional taxonomy or by numerical taxonomy based on the detailed description of characters. Whenever phenetic taxonomies are acceptable, automatic scanning and classification may provide a rapid and reliable approach. Problems of the size and orientation of the organisms remain to be worked out, but they should not present insuperable technical difficulties. The implications of the success of this method are that experience and insight into the presumed biological and phylogenetic significance of characters may be less important for obtaining satisfactory classifications than had been generally supposed.

Thus there is every reason to believe that classifications from automatically obtained characters are possible. This finding will, of course, lead not only to automatic classification but also to automatic identification, which should be one of the more exciting prospects for research workers faced with routine identification problems.

Numerical taxonomists working in biological taxonomy are continually surprised and impressed by the applicability of their principles in numerous sciences and other fields of human activity. They marvel at the rapidity with which this knowledge is spreading throughout the biological, medical, geological and social sciences, as well as the humanities. Numerical taxonomy has been employed to classify soils and diseases, politicians and plant communities, archaeological artifacts and oilbearing strata, socioeconomic neighborhoods and psychological types,

languages and television programs-to name just some of the applications. Sneath has even used it to solve a jigsaw puzzle. This broad spectrum of applications for numerical taxonomy should not surprise us. After all, the precise categorization of human experience is one of the foundations for a scientific understanding of the universe. We should not, however, be overly impressed by the similarities in approach in these various sciences. There are appreciable differences in the principles of classification in diverse fields, and it is necessary to know when the problems of one discipline part company with those of another. Nonetheless, the common fund of basic ideas on similarity and classification is great enough to serve as the basis for a general science of taxonomy.

Biological taxonomy will be affected by the computer in many ways besides numerical taxonomy. Automatic data processing will revolutionize the storage and retrieval of taxonomic information for museums and catalogues. The approaches of numerical taxonomy have already done much to de-emphasize the often legalistic and sterile aspects of naming organisms. It is likely that developments in automatic data processing will rapidly relegate problems of nomenclature to the position of relative unimportance they merit. Some of the birth pangs of automation will be felt in taxonomy as in other fields, and traditionally-minded workers will presumably resist the changes. The controversy about numerical taxonomy will doubtless continue for some time to come until a new "synthetic" theory of taxonomy, accepting what is soundest from various schools, becomes established. The revolution the computer has wrought in taxonomy has only just begun.



### Ionic conduction study leads to new concept in energy conversion device at Ford Motor Company.

Heart of the new sodium-sulfur battery recently announced by Ford Motor Company is a highly ionically conducting poly-crystalline ceramic, based on the compound beta alumina Na<sub>2</sub>O • 11A1<sub>2</sub>O<sub>3</sub>. This compound exhibits unusual two-dimensional diffusion and ion exchange properties.

A summary of the ion exchange properties of beta alumina when in equilibrium with mixtures of molten nitrates of monovalent cations is shown in Figure 1. The symbol X designates the cation shown with each curve.



Fig.1 Equilibria between beta alumina and various bimary nitrate melts containing NaNO<sub>3</sub> and another meltal nitrate at 300-350°C.

No anions of the molten salt enter the solid beta alunina phase. The equilibrium is unfavorable with cesium because the cesium ion is larger than the volume available at the cation location in the beta alumina, and is unfavorable for the lithium ion because this small ion can coordinate closer to the anions of melt than of the beta alumina.

In spite of the fact that beta alumina is a robust material with a melting point approaching 2000°C and a demonstrated resistance to molten salts, the diffusion rate of sodium ions in this material is as fast as the best solid inorganic ionic conductors known. For example, when a 2mm diameter single crystal of this material is placed in molten sodium nitrate at 300°C, it will exchange one half of its sodium ions with the sodium ions of the melt in 60 seconds. In more scientific terms, the diffusion coefficient for the sodium ion in the crystal is  $\sim 1 \times 10^{-5} \text{cm}^2/\text{sec}$  at 300°C and  $4.0 \times 10^{-7} \text{cm}^2/\text{sec}$  at 25°C.

The two-dimensional character of the diffusion is strikingly shown in Figure 2. The sodium ions of this thin single crystal of beta alumina were originally completely exchanged with the lithium isotope Li<sup>6</sup>.



Fig.2 Partially exchanged ( $\sim 60\%$ ) Li<sup>6</sup> single crystal of beta alumina. The area of high Li<sup>6</sup> concentration has become dark under neutron irradiation ( $\sim 10^{13}$  neutrons per cm<sup>2</sup> for 1 hour).

The crystal was then exchanged for a short period of time with molten  $\text{Li}^7\text{NO}_3$  which resulted in a replacement of part of the  $\text{Li}^6$  atoms with  $\text{Li}^7$  atoms. The crystal was then irradiated with neutrons in the Phoenix reactor of the University of Michigan which caused the area containing a large amount of  $\text{Li}^6$  atoms to turn dark in color. As can be seen in the photograph, the center portion contains the unexchanged  $\text{Li}^6$ . This is a direct visual indication that the exchange is two dimensional in the plane of this thin crystal.

Science is always interested in findings that indicate a potentially new form of energy conversion. The development of this high-energy secondary battery system is a significant step in that direction.

### **PROBING DEEPER FOR BETTER IDEAS**

