Toward a Perceptual Model of Folk Biological Classification

Eugene Hunn

Stable URL:

American Ethnologist is currently published by American Anthropological Association.

toward a perceptual model of folk biological classification[1]

EUGENE HUNN—*University of Washington*

Biologists engaged in field studies in exotic places have not infrequently noted that the local natives, who are entirely ignorant of Western European Science, nevertheless demonstrate an intimate and empirically reliable knowledge of the local flora and fauna and share with the fieldworker an appreciation of the ordered complexity of the living world. Diamond (1966) in New Guinea, Dennler (1939) in South America, and Irving (e.g., 1953) among the Eskimo have contributed to the anthropological literature by reporting detailed correspondences between the classes of organisms named by the natives and the taxa of scientists. Anthropologists have been slow in recognizing the strength of these similarities, perhaps preferring to accept the early verdict of Durkheim and Mauss: "Far, then, from man classifying spontaneously and by a sort of natural necessity, humanity in the beginning lacks the most indispensable conditions for the classificatory function" (1963:7).

However, recent advances in the anthropological understanding of folk biological classification stem in large part from a growing recognition of the striking similarities between folk systems and the biologists' system of classification. It is now clear that these similarities are neither fortuitous nor to be taken for granted. We are now directing our attention at discovering the constraints in nature and in the human mind which may account for the pan-cultural regularities (cf. Berlin, Breedlove, and Raven 1973).

(1) The vast majority of folk categories of plants and animals recognized in the systems so far described seem to be generated by primary attention to morphological and behavioral characteristics intrinsic to the organisms classified. Biosystematists employ a similar strategy in classifying living organisms: "Among a number of species some will resemble each other more than others. It is to be assumed that the resemblances, as far as homologous, reflect propinquity of descent" (Simpson 1961:116). Such categories are typically intractable to definition by reference to one or a small set of explicitly labeled attributes, but rather reflect perceptions of a totality of similarity and difference among

The notion of taxonomic structure has played a central role in recent descriptions and analyses of folk systems of biological classification. The increasingly apparent inadequacies of that notion as a model of folk classification processes justify a fundamental theoretical reorientation. Reasonably interpretable and formally adequate definitions of inductive classification (Postulate I) and of dissimilarities in a classification space (Postulate II) are more adequate than the taxonomic model for understanding patterns observed in folk biological classification systems. A non-rigorous extension of this "perceptual model" deals with the key problem of taxonomic ranks.

the organisms classified (cf. Bulmer 1970:1078; Hunn 1973:8-9). The established analytical techniques of ethnosemantics, such as componential analysis, have proven inappropriate for defining such concepts (cf. Turner 1974:16-17).

(2) A hierarchic structure of set inclusion relations among taxa is reported for all systems so far described. Though the depth of hierarchy found in folk systems nowhere approaches that of the scientific system, it may be noted that no folk systems deal with numbers of taxa comparable to those treated by the scientific system. Axiomatic treatments of this hierarchic property have been devised by Gregg (1954, 1967) for biosystematics and by Kay (1971) for the ethnographic context. Kay's *taxonomic structures* are a special case of Gregg's *Linnaean structures* (Kay 1971:885-886).
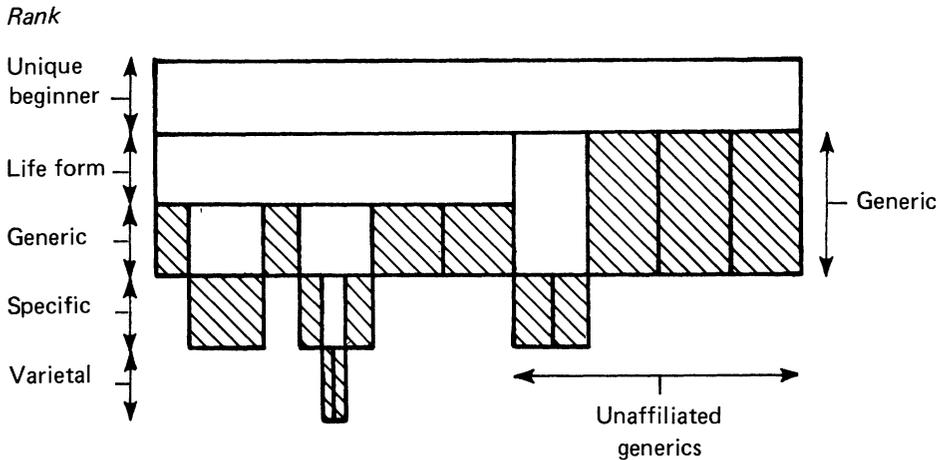
Rank

Figure 1. Block diagram of a taxonomic structure after Kay (1971), with folk taxonomic ranks following Berlin, et al. (1973).

*In Gregg's terms (1967) this is a truncated, irregular, and closed Linnaean structure. It is closed because each contrast set is a partition of the taxon which immediately includes the members of the set. It is only partially regular because the taxa of a given rank need not be immediately included in taxa of the next higher rank (e.g., the unaffiliated generics, but it is regular below the generic rank). And it is truncated because terminal taxa (crosshatched) need not be of the lowest rank.

(3) Essential to the biologists' classification scheme are categories (i.e., *ranks*) of taxa such as the Species, Genus, Family, etc. Defining taxonomic ranks makes possible the recognition of monotypic taxa so that we may appreciate the atypical singularity of a species such as the aardvark (*Orycteropus afer*) which occupies an order (Tubulidentata) all its own. Such taxa, to my knowledge, have not been reported in the ethnographic record; but Berlin (1973:260ff.) argues persuasively that folk taxonomic ranks must be defined in order to account for the structure of folk systems of classification and nomenclature. For example, binomial nomenclature is employed in folk systems much as it is in the scientific system. This implies that a contrast between taxa of *generic* and *specific* ranks is pervasive (Berlin, Breedlove, and Raven 1973:219-222). There is evidence for the existence of at least five discrete ranks of folk biological taxa.

Given the parallels between folk classification and biosystematics, it is not surprising that the primary conceptual device which has been used to describe and compare folk biosystematics is the *taxonomy*. Nevertheless, I believe that the limitations of the
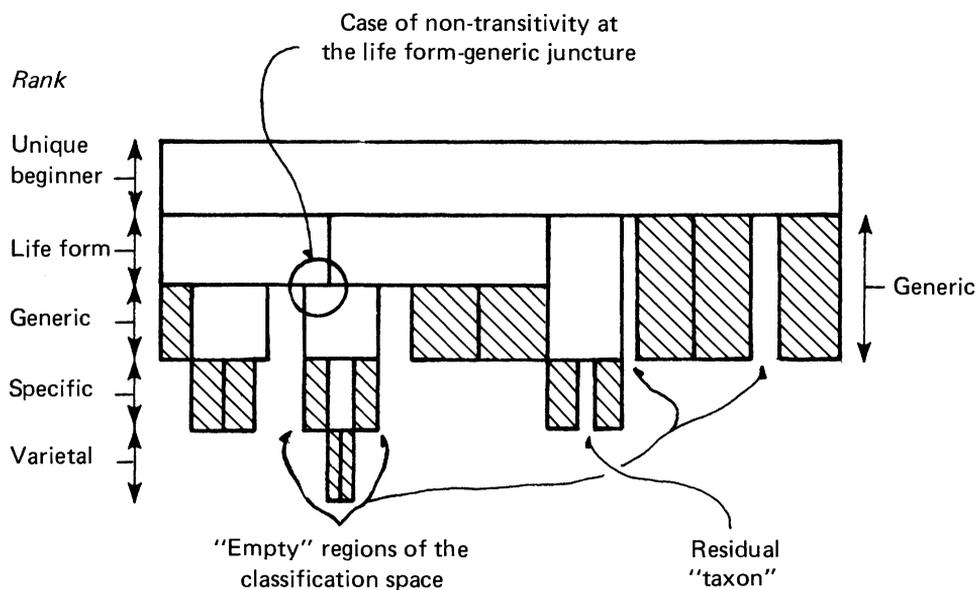
Figure 2. Block diagram of an imperfect quasi-taxonomic structure
illustrating empirical problems.*

*In Gregg's terms (1967), this is a truncated, irregular, but *open* Linnaean structure (ignoring the case of non-transitivity). The openness follows from the existence of unclassified elements at various ranks of the structure (some indicated by arrows). Life form taxa as well as some specific and varietal taxa will be deductive; generics and some subgenerics will be inductive. It can be seen that "residual taxa" and "empty regions of the classification space" are structurally equivalent phenomena. Residual taxa are simply empty regions below the generic level.

taxonomic model in the analysis of folk biological classification are already apparent. None of the following problems is adequately treated by reference to current taxonomic notions.

(1) Kay's axiomatic treatment of taxonomic structures simply postulates the existence of taxa. It does not explain how the taxa are generated nor why a particular organism is placed in a particular taxon. The key fact that the vast majority of folk taxa correspond to scientific taxa (Hunn 1975a) can only be explained by reference to common processes of concept formation applied to a shared world of objects. As noted above, traditional techniques of semantic analysis, which attempt to explain concepts by specifying necessary and sufficient conditions for membership in the concepts, have proven inadequate for defining the basic folk taxa, which are semantic primitives.

(2) The recognition of taxonomic ranks, which seem to play essential roles in both scientific and folk systems, requires inelegant complications of formal taxonomic models (Kay 1971:886). Such ranks also pose methodological puzzles. Berlin's important claim that taxa of the generic rank—taxa said to be "psychologically basic" and "semantic primitives" (1972:54-55)—will be named before any more or less inclusive taxa rests on the artful weighting of linguistic, formal, and biological considerations necessary to determine which taxa are generic (Hunn 1973:88-92). Thus this key aspect of recent theoretical work is not adequately grounded in current taxonomic theory.

(3) The rarity of named suprageneric taxa has been noted by most investigators. D'Andrade (cf. Berlin, Breedlove, and Raven 1968:298) at an early date called attention to this fact as requiring some explanation. The result was the development of techniques for recognizing covert taxa (Berlin, Breedlove, and Raven 1968; Hays n.d.; Brown 1974),

most of which subsume two or more generic taxa (Berlin, Breedlove, and Raven 1974:36).

Though Kay's formal definition of taxonomy does not preclude the existence of covert taxa (1971:881), such taxa pose numerous problems for a taxonomic theory. How can the boundaries of covert taxa be defined with the clarity required by the taxonomic model? My Tzeltal zoological data indicate that many of the covert mid-level groupings are best interpreted as *chains*, often with indeterminate boundaries. Chains are groupings of taxa constructed by reference to "horizontal" relations among members of a single contrast set, relations like similarity and difference (Hunn 1975b). Unlike the relation of set inclusion, which characterizes taxonomic structures, the similarity-difference relation is non-transitive, and a system of similarity-difference relations will differ in fundamental ways from systems generated by transitive relations. The taxonomic model does not take non-transitive horizontal relationships into account.

Non-transitive relationships are important in many folk systems. For example, the Brights report that it is difficult to accommodate certain features of Northwestern California Indian folk zoology with the taxonomic framework. They conclude

> that a hierarchical model, which shows only the relationship of *domination* ('A dominates B' = 'B is an A'), cannot account adequately for the Indian taxonomies . . . there is no way of indicating when an item is classified in a certain way because it is 'like' another item which is more central to the focus of the domain in question (Bright and Bright 1965:253).

They introduce a "sphere of influence model" to account for the fact that certain plants are classed as "bushes" because they are "like the lilac" which is a kind of "bush" (see Figure 3). They note that this model is only "partly hierarchical." Turner cites numerous examples of similar ambiguity in her analysis of suprageneric folk taxa in three Northwest Coast Indian languages (1974:31-47).

Berlin, Breedlove, and Raven encounter precisely this situation with the Tzeltal. Many plants are not considered to be "true" examples of any recognized generic taxon but are recognized and classified as "almost the same as" some well known generic. Their solution involves distinguishing "basic" and "extended" ranges within the designatum of a term (1974:56-58). In each case cited above, the structural problem is beyond the scope of the taxonomic model, requiring reference to horizontal relations of similarity-difference in addition to the vertical taxonomic relation of set inclusion.

(4) My recent work on Tzeltal folk zoology (1973, 1975) has revealed additional structural phenomena which cannot be accounted for within the taxonomic framework. *Residual taxa*, like the Tzeltal generic taxon, *pehpen* 'butterfly/moth' (Hunn 1973:120-123; see Figures 4 and 7), do not fit the taxonomic model. *Pehpen* includes six consistently recognized specific taxa such as *me?saƛ' pehpen* 'mother-of-the-sphinx-moth butterfly.' After these particularly noteworthy types (from the Tzeltal point of view) have been delineated, the remainder are treated as undifferentiated *pehpen*, 'just butterfly,' or are dealt with by applying a descriptive phrase which modifies the generic name. In strictly taxonomic terms, residual taxa must be treated as specific taxa. But the residual subdivision of *pehpen* is logically quite different from most taxa, i.e., it is defined by the *absence* of any distinctive perceptual unity. Residual taxa also occur in English, e.g., a "mutt" is a "dog" which is not otherwise distinguished.

A closely related problem follows from the existence of *unlabeled regions of "taxonomic space."* Certain uncommon and inconsequential animals are recognized by the Tzeltal only in terms of their life-form or domain application (Hunn:116-119), much as we might recognize a plant as "some kind of tree." Turner describes precisely this situation in three Northwest Coast Indian languages, noting the existence of " 'empty'
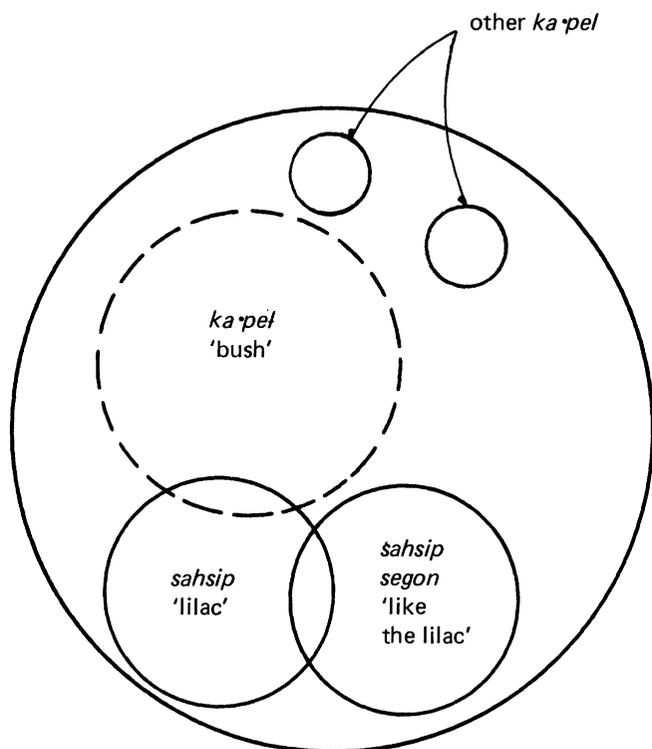
Figure 3. Bright and Bright's "sphere of influence" model: Yurok "bush."*

This diagram (adapted from Bright and Bright 1965:254) is intended to emphasize the horizontal relationships expressed by the term "like." The "lilac" category is seen as similar to the core (broken circle) of the "bush" category, and the category "like the lilac" falls in the broad range of "bush" (largest circle) *because* it is linked through the "lilac" category by the horizontal relation of similarity.

life-form taxa" (1974:35, 69). Since taxonomic structures *sensu* Kay require that the taxa of a contrast set *partition* their common immediately superordinate taxon, the indeterminacy described above requires *ad hoc* adjustments.

(5) The related problems of *non-transitivity*, raised by Kay (1975) and Randall (1976), and of *crosscutting classificatory principles* (cf. Black 1967; Perchonock and Werner 1969; Brown 1974) seem to require a revision of the theoretical basis of our descriptions of folk biological classification. The problem of non-transitivity—a problem for the analyst, not for the folk—has been most often noted at the juncture of life form and generic taxa, e.g., "willows" are a kind of tree (many English speakers would agree), but some individual willows or certain kinds of willows may be "bushes" (Randall 1976). This violates the taxonomic premise that if A is included in B and B in C, then A is included in C (i.e., transitivity of the relation of set inclusion). Crosscutting classifica-. tions, if not clearly isolated, may also violate this premise.

This list of unsolved problems is not exhaustive, but it is sufficient to motivate a reconsideration of the current theoretical framework.

beyond taxonomy: outline for a perceptual model

In defining the following alternative model of folk biological classification, I have
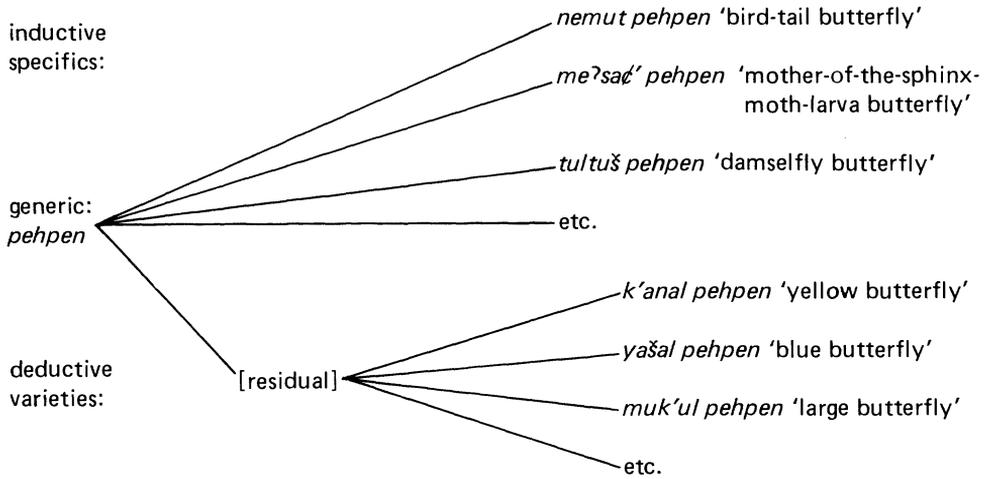
Figure 4. A taxonomic representation of a covert residual taxon:
the Tzeltal taxon, *pehpen* 'butterfly/moth.'*

Most adult Lepidoptera are called *pehpen*. Binomial nomenclature is liberally used to distinguish kinds of *pehpen*. The attributes labeled in most binomial terms are interpreted literally and applied to scientifically arbitrary subsets of the Lepidoptera. A few of these terms, however, are consistently applied to scientifically recognized subdivisions of the Lepidoptera. In these cases the attributes labeled are not interpreted literally. In the first instance, the "names" distinguish deductive varieties. In the second, they distinguish inductive specific taxa. Furthermore, "names" of the first type are restricted in their application to a residual subdivision of *pehpen*, i.e., all *pehpen* not labeled by names of the second type.

relied heavily on insights from cognitive psychology and biosystematics. The relevance of these fields for our problem is obvious. If we rework a discussion by Bruner, Goodnow, and Austin (1956) about the role of attributes in concept formation, we may recognize three possible paths to the construction of a concept by reference to descriptive attributes (see Figure 5). First, a set of defining attributes may be specified such that an element is an example of the concept if and only if the description of the element with respect to each attribute falls within a defining range of values. For example, an object is a "red ball" if and only if it is perceived to be red in color and spherical in shape. Concepts of this type require that the defining attributes be few in number, because as Bruner, et al. suggest, an increase in the number of defining attributes means that energy conserving simplifications will be employed (1956:47). In addition, we expect these defining attributes to have broad relevance. In other words, color and shape are attributes relevant to an enormous catalogue of things. It is thus natural that particular colors and shapes are themselves named. The values of widely applicable attributes will be referred to as *abstract features*.

Bruner, et al. (1956:47) cite two techniques for reducing the mental energy required in utilizing multi-attribute concepts. The first is *attribute reduction*. This procedure involves a selection of a limited set of defining attributes from the set of criterial attributes. Attribute reduction results also in concepts defined with reference to abstract features.

The second technique is *configurational recoding*. According to Bruner, et al.

Such reconstruction is possible because in fact the defining features of most objects and events are redundant with respect to each other. A bird has wings and bill and feathers and characteristic legs. But . . . if it has wings and feathers, the bill and legs are highly predictable.
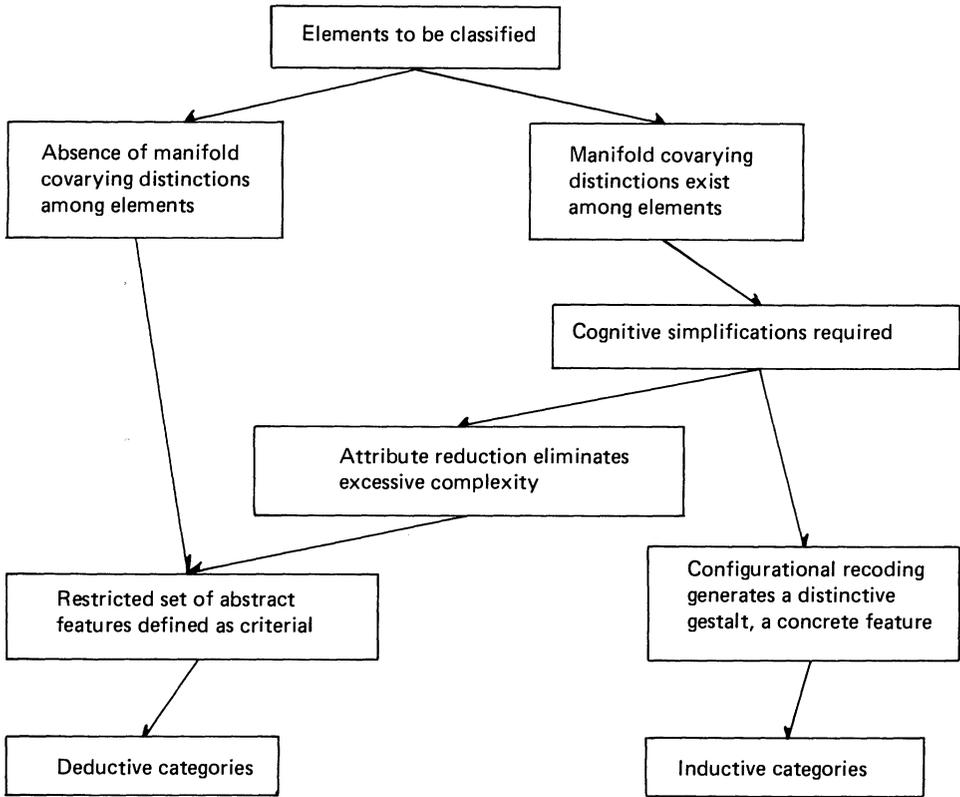
Figure 5. Varieties of categorizing strategies.

> In coding or categorizing the environment, one builds up an expectancy of all of these features being present together. It is this unitary conception that has the configurational or Gestalt property of 'bird-ness' .... When the conception is well enough established, it takes on the property of being able to serve as a discriminable and seemingly irreducible attribute of its own (1956:47).

Thus configurational concepts, unlike the other types discussed above, are recognized by reference to a *single* attribute rather than by reference to the concatenation of several attributes. Nevertheless, the recoding which yields a configurational attribute is made necessary by the recognition of manifold differences between sets of objects or events. We may refer to such unique configurations as *concrete features*.

The distinction between concepts defined by abstract features and those defined by concrete features may be rephrased in terms of the two fundamental logical processes, deduction and induction. Deduction generates categories *a priori* by the statement of a defining rule. The features referred to in the defining rule will be abstract features, since an *a priori* definition must be phrased in terms of attributes which have been defined prior to and independent of their use in the definition of the concept in question. Inductive concepts do not depend on *a priori* definitions, but rather may be constructed by abstracting configurations from a sample of as yet unclassified objects. No defining rule need be given. In fact, as a rule, explicit definitions of inductive concepts *cannot* be stated, since the configurations with reference to which such concepts are constructed are *unique* to each concept. In English we conventionally refer to these concrete features by suffixing "-ness" to the name for the concept. Thus "raccoonness" is the configuration

by which members of the category "raccoon" are recognized.

The first postulate of the model of folk biological classification proposed here asserts that (I) *folk biological taxa are generated by induction*. I do not deny the existence of deductive classifications; I only argue that deductive classifications are not necessarily taxonomic (more precisely, they will not generally fit the definition of Linnaean Structures; [Gregg 1967]), but that inductive classifications will be. Distinguishing these two types of classification makes it possible to account for problems like "non-transitivity" (see below).

A second postulate is suggested by distinctions made in biosystematics. Simpson paraphrases criteria for recognizing supraspecific taxa (species taxa being defined, ideally, by reference to population dynamics) suggested by Mayr, Linsley, and Usinger (1953): "A T_{j+1} is a taxon including one or more T_js of inferred common phylogenetic origin and *separated from other T_{j+1}s by a decided gap*" (italics mine; Simpson 1961:191). The presence of a "decided gap" or discontinuity with reference to multiple attributes permits the creation of a "non-arbitrary" group with respect to exclusion (Simpson 1961:115-116). The recognition of "decided gaps" implies that degrees of divergence within a set of taxa are compared, for a gap is judged as "decided" only in comparison with the extent of other gaps. These considerations imply that biosystematists think in terms of a "classification space" in which "distances" within and between taxa are evaluated and compared in terms of overall similarity and difference.

I believe that the observed structure of folk biological classifications—both that which accords with the taxonomic model and that which does not—may be largely accounted for in terms of perceptions of relative widths of gaps between taxa and of degrees of heterogeneity within taxa. Thus my second postulate asserts that (II) *human beings perceive organisms as existing in a "classification space." "Distances" within this space are assessed in terms of overall similarity and difference between organisms with respect to perceptible attributes of morphology and behavior. Taxa are defined by reference to patterns perceived within this system of differences.*

the perceptual model

We may now consider how these postulates might be used to define the concepts and relations needed to describe the empirical regularities noted in folk systems of biological classification. Figure 6 is a simplified graphic representation of two possible perceptions of the distribution of organisms in the classification space. In the first instance (A), organisms are distributed more or less randomly in the sector of the space shown. Groups or clusters are not readily apparent. In the second (B), the clusters are obvious. I contend that biological organisms will typically exhibit obvious clustering as in Figure 6B. With the exception of polymorphisms or domestication, evolutionary and ecological theory leads us to expect that all individuals of a biological species in a limited geographical range are capable of being perceived as forming an obvious cluster in the classification space.[2] Residual taxa are anomalous because they do not depend on the perception of such a pattern (see Figure 7). The perception of clusters at higher levels is obviously more problematic.

The ideas of a "classification space" and of relative "distances" between percepts are understandable as *metaphors*. However, it is necessary at this point to make these metaphors explicit. Algebraic treatment is a means to this end. It is essential that an algebraic treatment be "reasonable" as well as precise. I have not yet succeeded in discovering a model which is both formally adequate and reasonably interpretable that

A. Exhibiting no obvious pattern, thus pretaxa are not defined.
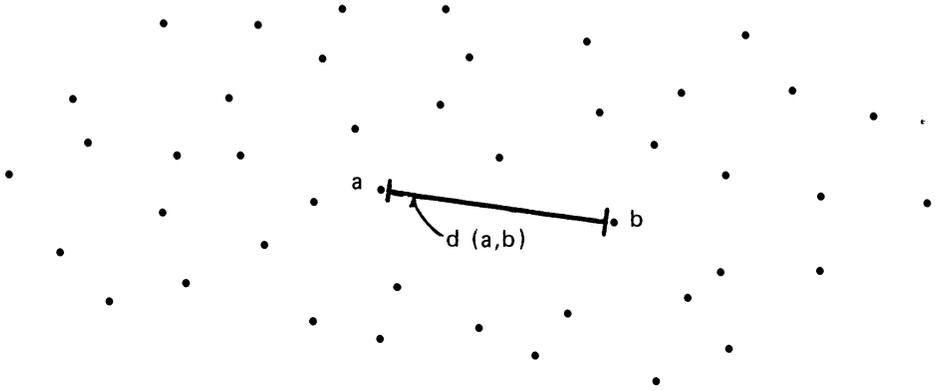
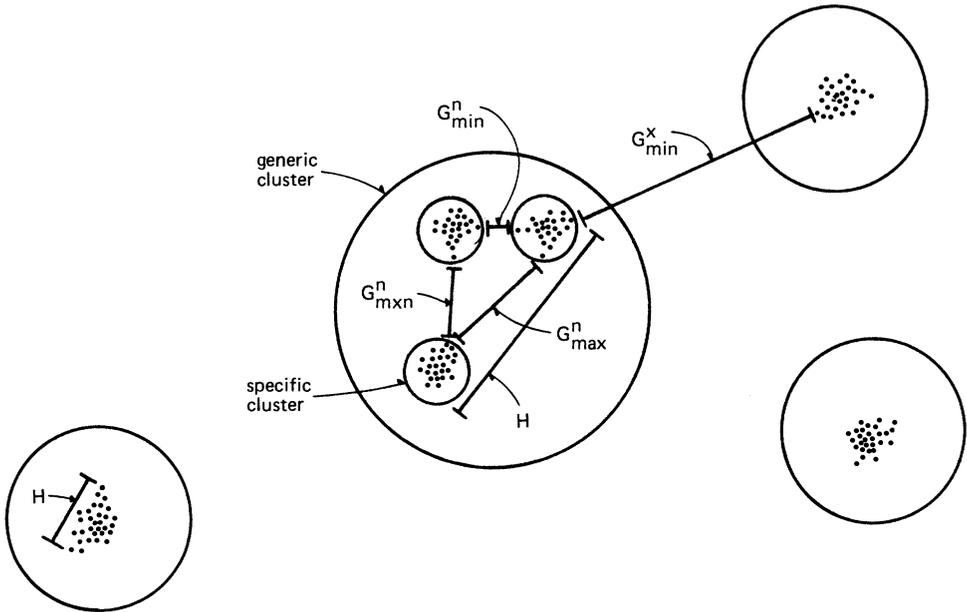B. With obvious clustering evident at two levels of inclusion.

Figure 6. Percepts in the segment of a classification space.*

*The present two-dimentional representation is not intended to imply that the classification space has a determinate number of dimentions. In fact, the concept of "dimention" is essentially metaphorical here. A linear format is also useful as a heuristic device (cf. Hunn 1973:100 ff).

e.g., *tultuš*
'dragonfly'

pehpen

A pehpen

B pehpen

C pehpen

residual *pehpen:*
to include *X pehpen,*
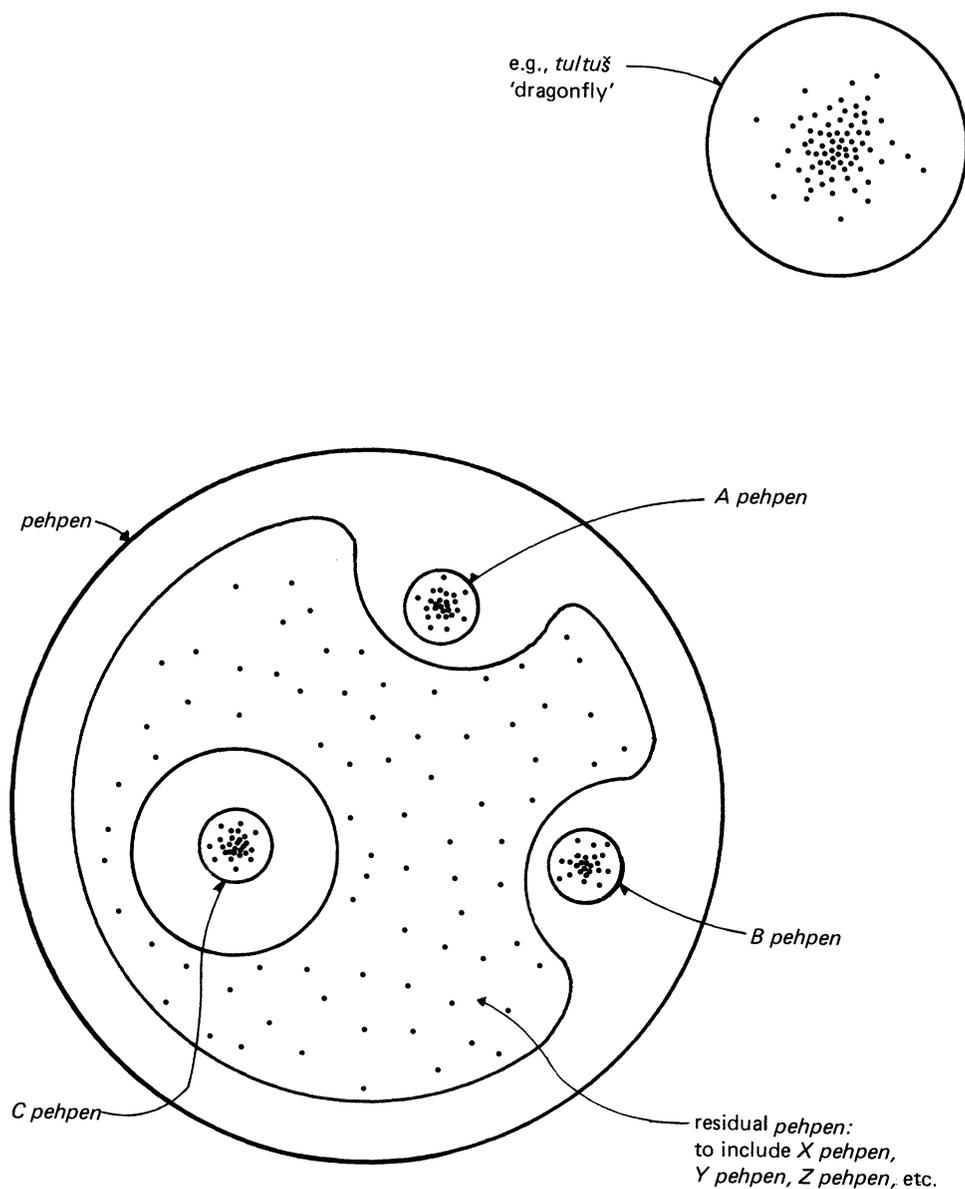*Y pehpen, Z pehpen,* etc.

Figure 7. A spatial representation of the Tzeltal generic taxon, *pehpen*,
illustrating the anomalous structure of a residual specific taxon.'*

*The attributives *X*, *Y*, and *Z* are literally descriptive, thus *X-pehpen*, etc., are descriptive phrases, not names, and the categories so labeled are deductive. It can be appreciated that the residual (and any deductive subdivision of it) does not qualify as a pretaxon; there are distances within the residual greater than distances from points in the residual to certain external points.

will succeed wherever the taxonomic model has failed. Thus I outline below (1) a formally adequate and reasonable model capable of accounting for some, but not all, of the descriptive problems cited above (see the Appendix for a full explication), and (2) a descriptively adequate model which is either precise or reasonable but not both. Both

models are "perceptual" models in that they define an algorithm for generating a system of categories from perceived differences between percepts.

the formally adequate but descriptively limited model   Let U be the set of all "percepts" of organisms in an individual's "classification space." Let us assume that each pair of percepts may be compared with respect to overall similarity as asserted in our description of inductive classification above. If x and y are any two percepts in U, then $d(x,y)$ may stand for the difference between them, or simply their perceived degree of dissimilarity. The "classification space" may be defined, for the present, as a partially ordered set (a *poset*), the elements of which are *not* percepts but the perceived differences between pairs of percepts. This incorporates the assumption that people are able to judge, in certain cases, whether the difference between percepts a and b is greater than or equal to the difference between percepts c and d. Further, we define the difference between a percept and itself (i.e., $d(x,x)$) as equal to zero and difference as symmetrical (i.e., $d(x,y) = d(y,x)$). A critical assumption adds structure to this model of folk biological classification: i.e., the differences from a *given point* to all other points are directly comparable and thus completely ordered by the greater-than-or-equal-to relation, (i.e., $d(x,y) \leqslant d(x,z)$ or $d(x,z) \leqslant d(x,y)$, all $x,y, z \in U$). This assumption seems reasonable and is necessary to what follows.

The simple structure so far defined allows us to define a minimal prerequisite for the inductive recognition of a taxon, i.e., a *pretaxon*. If any two points x and y are in a pretaxon, and if the distance between x and any third point z is not greater than the distance between x and y (i.e., $d(x,z) \leqslant d(x,y)$), then z is also in that pretaxon. With these assumptions it is possible to prove that the set of all pretaxa definable in U is almost but not quite a taxonomic structure as defined by Kay (1971). The axioms of the taxonomic model, with one exception, are demonstrable theorems in the perceptual model. Specifically it can be demonstrated that (1) any two pretaxa are either non-overlapping, or one strictly includes the other, and (2) U is the unique beginner (see Appendix). However, it is not necessarily the case that the set of all pretaxa immediately included in a given pretaxon is a partition of the superordinate pretaxon. It is noteworthy that Gregg (1967) specifically provided for this possibility. The ethnographic data suggest that the partition requirement of Kay's formulation is too strong, i.e., residual taxa and empty regions of the classification space violate this stricture (see Figure 2). The perceptual model can account for these taxonomic anomalies by reference to the definition of a pretaxon. Residual taxa and empty classification regions expectably occur where this requirement is not met. The Tzeltal classification of shorebirds is illustrative. Numerous species of this group (suborder Charadrii) occur rarely in the vicinity of the Tzeltal community, but the pattern of occurrence of individual birds is sporadic; thus for a member of this community the shorebird *percepts* are also scattered—they do not form obvious observable clusters. Consequently, there are no pretaxa in that region of the classification space. The result is a total lack of consistency in applying a set of "loose names" deemed appropriate for that general type of bird (Hunn 1973:166ff.).

The present formalism also suggests an explanation for the vexing theoretical problem of non-transitivity.[3] The proposed minimal criterion for the existence of a taxon (i.e., the definition of a pretaxon) is in context a definition of the inductive process of category formation. Since this process will not generate crosscutting categories, a system of inductive taxa should exhibit transitivity with respect to relations of set inclusion. However, life form and other suprageneric taxa are not, as a rule, generated by induction. For example, higher order plant taxa such as "tree," "bush," and "vine"—which occur in

English folk botanical systems as well as in those of many unrelated languages—are defined by reference to a small set of *abstract features*, primarily stem habit and leaf form (Berlin 1973:260-261; Berlin, Breedlove, and Raven 1974:30). This is true as well for "bird" and "four-legged quadruped," perhaps the most widespread zoological life forms. Unlike generic taxa, these are not defined by reference to *overall* similarity and difference. In short, life forms tend to be *deductive* categories, despite their morphological and behavioral basis. If an inductive taxon is subdivided deductively, transitivity is maintained, but if a deductive category includes inductive taxa, there is no guarantee that the inductive taxa will be included in their entirety. For example, American English speakers may be aware of the existence of "flying squirrels" (*Glaucomys* spp.). Given a suprageneric folk concept based deductively on mode of locomotion, this squirrel would almost certainly be considered a "flying animal" (actually, it can only glide), while other squirrels would be excluded. This "mode of locomotion" division crosscuts the *inductive* lower order taxon "squirrel." (Try to define "squirrel" in terms of a few abstract features.) It does so, predictably, because the higher order grouping is deductive, the lower inductive. Taxonomic theory provides no basis for distinguishing these two discordant processes of category formation, as my model does.

One critical problem remains unsolved in the present format; that is, taxonomic ranks remain *ad hoc*. The following formulation, though formally inadequate, may indicate a fruitful direction for future theoretical work.

the ratio of distances model   The above formalization defined the classification space by reference to a partial ordering of *relative* differences between pairs of points. No claim was made that *ratios* of differences were definable, since a partial ordering gives no basis for interpreting sums, products, or ratios. Let us now, for the sake of argument, define differences in our classification space as real numbers.[4] We are assuming that people have "tape measures" in their heads for measuring inter-percept "distances." This may be too strong an assumption, but it produces some interesting results.

Let G stand for the width of a *gap* between two taxa (clusters of organisms). G can be defined as the shortest distance between any member (i.e., a percept) of the first cluster and any member of the second (see Figure 6B). Let $H(t)$ stand for the *heterogeneity* of a taxon t. $H(t)$ may be defined as the largest distance between any two member percepts of the taxon t, or the distance "across the taxon." Superscripts allow us to distinguish gaps in two structural positions. $G^n(t)$ is the width of an *internal* gap, a gap between subclusters within the taxon t. $G^x(t)$ is the width of a gap *external* to the taxon t. Subscripts may then specify (1) the smallest gap of a set, G_{min} or (2) the largest, G_{max}. By combining subscripts with superscripts we can refer to several important types of gaps. For example, $G^n_{max}(t)$ refers to the *largest* gap between subclusters of the taxon t. If t were a polytypic generic taxon, $G^n_{max}(t)$ would be the distance between the two most dissimilar specific taxa included within it (see Figure 6).

We may now define an index of the *perceptual salience* of a taxon, that is, the degree to which a cluster of organisms is obvious. This is the *critical ratio*, CR(t), of a taxon t and is defined so that as $G^x_{min}(t)$ (the minimal external gap) increases, the critical ratio will increase, while as $H(t)$ (heterogeneity) increases, the critical ratio will decrease.[5]

It is now possible to propose a natural definition of the *generic taxonomic rank*: A taxon, t, is a generic taxon if and only if CR(t) is greater than CR(t ± n), where "±n" indicates a taxon which either includes the taxon t or is included within it. Thus, by definition, generic taxa are perceptually the most salient taxa of a series of nested taxa, as asserted by Berlin, et al. (1973:216).

This definition has interesting implications. For example, it is possible to define a situation in which the same taxon is both a generic and a life form, or, conversely, where a single taxon immediately includes both generic and specific taxa (i.e., if CR(x) < CR(y) < CR(z), x, z ⊂ y). I believe that this is a reasonable interpretation of several apparently exceptional cases cited by Berlin, et al. For example, *čikinib* 'live oak' (or *čikinib hihte?* 'armadillo-eared oak') has some characteristics of both generic and specific taxa (1973:224-225). Most informants include it together with several obviously specific taxa in the higher order taxon *hihte?* 'oak' (see Figures 8 and 9). Yet it is typically labeled by a primary lexeme, otherwise indicative of generic status. In Berlin, et al.'s scheme, *hihte?* must be an "intermediate" taxon (1973:225-226), if *čikinib* is a generic, and a polysemously labeled taxon, *(baḍ'il) hihte?* 'true oak,' must be postulated to include the remaining kinds of oaks, which are clearly specifics. Otherwise, the generic category would not partition the unique beginner, as required by Kay's formal treatment. Such borderline cases pose logical problems within the taxonomic framework, but they make good sense from the present perspective.

It could be argued that a system which allows a taxon to belong to more than one rank is defective. It is not defective; it is simply not taxonomic *sensu* Kay. Interpreting the Tzeltal taxon *hihte?* 'oak' as immediately including both generic and specific taxa in accord with the present model of folk classification eliminates an "exception" to the rule. I do not argue that the notion of taxonomic rank be abandoned, but rather that it be freed from the requirements imposed by an inappropriate logic.

conclusion

Taxonomic-like structures have proved convenient forms by which to organize human knowledge of variation among living things, for the biosystematist as well as for the student of folk classification. However, as models of the *process* of classification they are clearly inadequate; the existence of taxa is accepted as a primitive notion of these logical
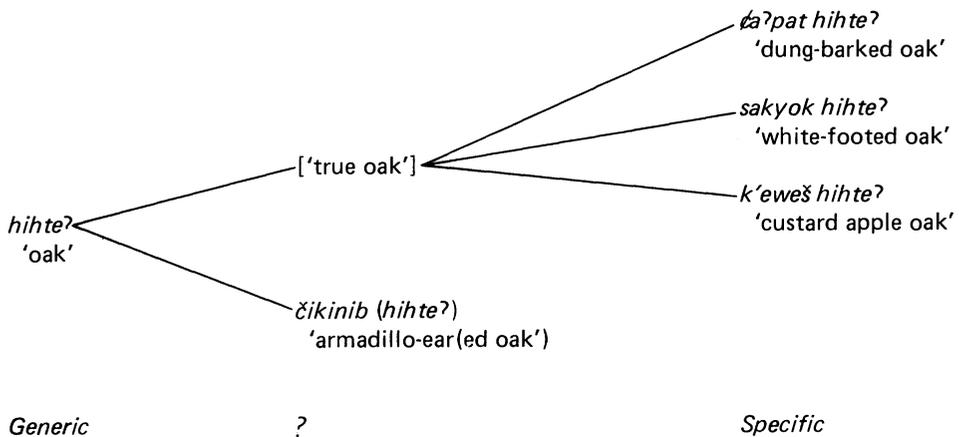

Figure 8. A taxonomic representation of the Tzeltal taxon, *hihte?* 'oak'
(adapted from Berlin, et al. 1973:225)*

*'True oak' might be postulated as a covert generic taxon and *hihte?* then classed as of "intermediate" rank. This would avoid leaving the rank of *čikinib* 'live oak' undefined. Berlin, et al. suggest that *čikinib* may be in the process of becoming a generic taxon. An alternative explanation is suggested in Figure 9.
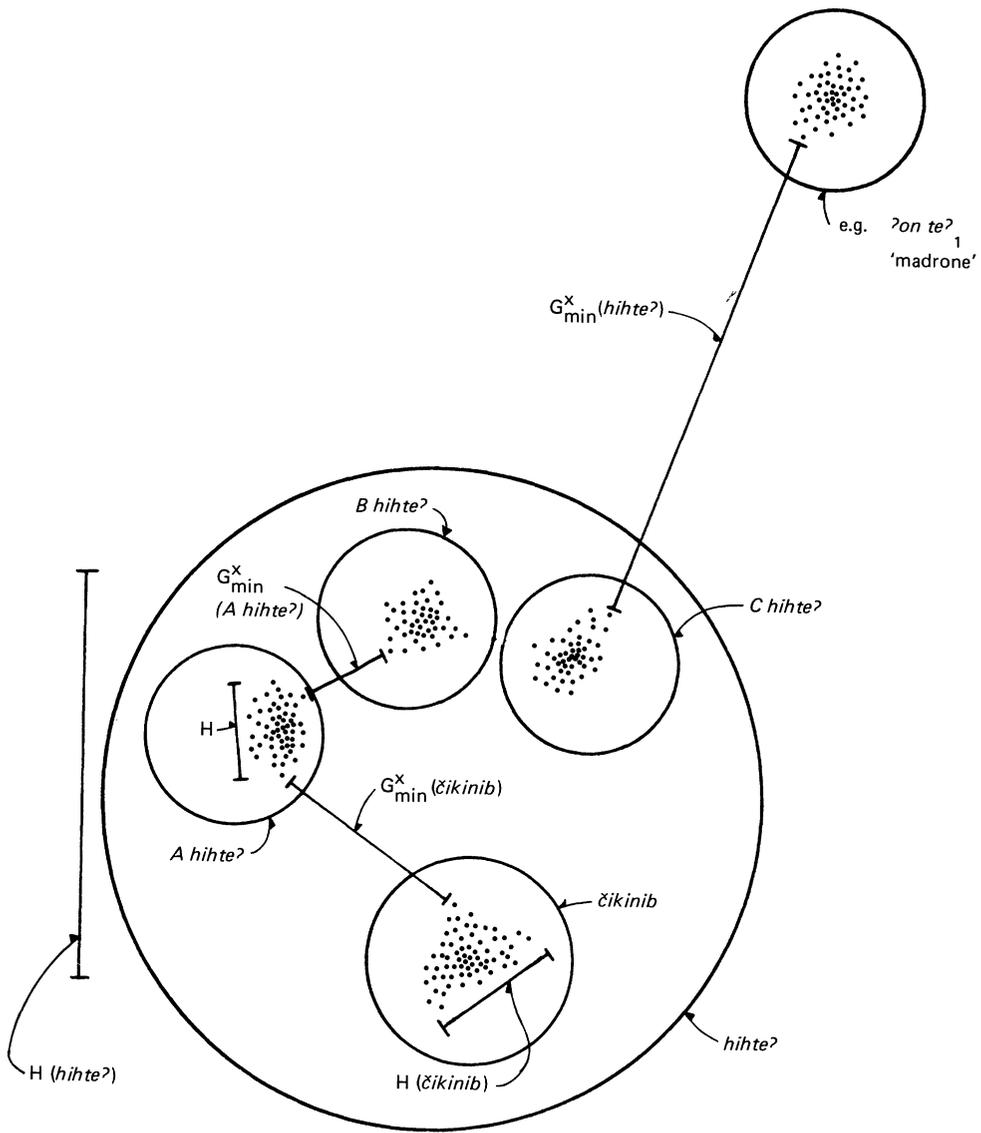
e.g. $\text{?}on\ te\text{?}_1$
'madrone'

$G^X_{min}(hihte\text{?})$

$B\ hihte\text{?}$

G^X_{min}
$(A\ hihte\text{?})$

$C\ hihte\text{?}$

H

$G^X_{min}(\check{c}ikinib)$

$\check{c}ikinib$

$A\ hihte\text{?}$

hihte?

H (hihte?)

H ($\check{c}ikinib$)

Figure 9. A spatial representation of the Tzeltal generic taxon *hihte?*
illustrating a situation of indeterminate rank assignment (from Berlin, et at. 1973:225).*

By the definition proposed here, if CR(cikinib)*>CR(*hihte?*), *hihte?* would be suprageneric with respect to *cikinib.* At the same time, CR(*hihte?*) may be greater than CR(*A hihte?*), etc., in which case it is generic *with respect to* sub-taxa A, B, and C.

structures, which are then constructed by reference to the relation of set inclusion. The fundamental problem, however, is to account for the *existence* of taxa. To do this requires that both the taxa and the taxonomic-like structures be shown to come from a more fundamental logic. I have proposed here that *percepts* be taken as the starting point and that a variable *distance* (meant to represent perceived degrees of overall similarity-difference) between pairs of percepts be employed as the fundamental relation. This

a perceptual model   521

strategy shows promise of generating near-taxonomic structures which lack none of the advantages of such structures but which conform more closely to patterns actually observed in the data. Though no doubt taxonomic-like forms will continue to be useful for representing observed data patterns, these forms may now be seen to represent the end product of a clearly non-taxonomic cognitive process.

appendix: explication of the formally adequate model[6]

First, define the universe of basic elements as the countable set U, the elements of which may be thought of as the "percepts" of individual organisms seen by an individual folk biological classifier. Accept the fiction that the classifier sees all these percepts simultaneously rather than over a period of time.

Next, define a structure, $(P, \leqslant)$, which is a partially ordered set (a *poset*) with a zero element ($\emptyset \in P$). This structure may be used to define an individual's *classification space*.

Assume that a classifier can compare any two percepts. Then the *difference* between two percepts, their dissimilarity, is modeled by a function d, which maps each element of the power set $U \times U$ onto the poset P,

(1) $d: U \times U \longrightarrow P$ (onto).

Thus $d(x,y)$ stands for the perceived dissimilarity of the percepts x and y, for all x and y in U. (Note that the ordered pair (x, y) is an element of the power set $U \times U$.) The mapping onto the poset P implies, further, that a classifier may compare the dissimilarities of pairs of percept pairs under certain circumstances.

It is reasonable to assume further that in case

(2) $d(x,y) = \emptyset$ then $x = y$.

In other words, the dissimilarity between a percept and itself is zero (the zero element of the poset P, as specified above). All other differences are positive. In addition let

(3) $d(x,y) = d(y,x)$.

That is, dissimilarity is symmetrical.

Now it is necessary to specify the conditions under which the differences between pairs of percept pairs may be compared. First, define the *chain of x in A*, C_x^A, for all A, $A \subseteq U$, and for all x, $x \in A$,

(4) $C_x^A = \left\{ d(x,y) \mid y \in A \right\}$.

Assume that all such chains are *completely ordered sets*, i.e., for all x, y, and z in A,

(5) $d(x,y) \leqslant d(x,z)$ or $d(x,z) \leqslant d(x,y)$.

In other words, a classifier is deemed capable of deciding which pair of percept pairs is most different whenever the two pairs share a common percept.

Now define a *pretaxon*—intepreted as a necessary but not sufficient condition for a subset of U to be an inductive taxon—as follows:

(6)   A, a subset of U, is a *pretaxon* in case
(a) A is neither null nor a singleton set,
(b) for all x and y in A and for all z in U if $d(x,z) \leqslant d(x,y)$, then $z \in A$.

In short, this definition embodies the claim that all inductive taxa must be pretaxa. Thus the dissimilarity between any member of such a taxon and any other member must be less than the dissimilarity between that member and any element that is excluded. This defines the process of induction by specifying for each subset of U whether or not that subset can be an inductive taxon.

Given the definitions and assumptions specified above, it is possible to demonstrate that the *essential* properties of taxonomic structures necessarily follow. According to Kay (1971:881), a taxonomic structure has the following properties.

(1) Def. A binary relational structure $\tau = \langle T, \supset \rangle$, where T is a finite non-null set of non-null sets and $\supset$ is the relation of strict inclusion restricted to the members of T, is a TAXONOMIC STRUCTURE iff ('if and only if') the following two axioms are satisfied. 2) Axiom. There is a unique $t_1 \in T$ called the UNIQUE BEGINNER, such that, for any $t_i \in T$ ($t_i \neq t_1$), $t_1 \supset t_i$. 3) Axiom. For any $t_i \in T$, the [contrast] set $c(t_i) = \left\{ t_j \mid t_j \in T, t_i \supset t_j \text{ and there is no } t_k \in T \text{ such that } t_j \supset t_k \text{ and } t_k \supset t_j \right.$.

If we equate Kay's set of taxa T with the set of all pretaxa as defined here, it is obvious that the set of all pretaxa is finite—since U is countable—and non-null—since U is trivially a pretaxon and, incidentally, the unique beginner required by Kay's first axiom (2). It should also be clear that by definition 5a above, all pretaxa are non-null also, as required.

It can be shown that for all pairs of pretaxa, either their intersection is null or one strictly includes the other. Thus the relation of strict inclusion, $\supset$, defines the relations among all pretaxa of a domain.

(7)   Claim: for all subsets A and B of U,
            if A and B are pretaxa,
        $A \cap B = \emptyset$ or $A \cap B = A$ or $A \cap B = B$.

To demonstrate this, assume the contrary, that is, assume that there is an $x \in A \cap B$, an $a \in A \cap B'$ (B' means "not B"), and a $b \in B \cap A'$. Thus $d(x,a) < d(x,b)$ since x and a are members of A and b is not. Likewise $d(x,b) < d(x,a)$ since x and b are members of B and a is not. This implies the contradiction, $d(x,a) < d(x,a)$. Thus $A \cap B = \emptyset$ if a and b exist.

However, Kay's second axiom (3), i.e., that the members of a contrast set partition the set in which they are immediately included, does not necessarily follow. There is no formal impediment to the existence of a percept x in the inductive taxon A such that A strictly includes one or more inductive taxa P_i and x is *not* a member of any P_i, $P_i \subset A$. Such an x will exist under the following conditions. For all pretaxa $P_i \subset A \subseteq U$, and for all $r, s \in P_i$, if $d(r,s) < d(r,x)$, and for all $u \in A$, some $v \in A$, and all $z \in A'$, if $d(u,x) \leqslant d(u,v) < d(x,z)$, then $x \in A$, $x \notin P_i$, as required.

notes

[2] The "competitive exclusion principle" of biological ecology suggests that two coexisting species cannot occupy the same "niche." It is inevitable that differences in niche will be reflected morphologically and/or behaviorally. Evolutionary theory defines species as breeding isolates. This also implies that two species, especially coexisting species, will differ phenotypically to a significant degree. Given an essential minimal difference between any two coexisting species, it is not unreasonable to assume that those differences could support the inductive recognition of species as folk taxa. I would like to thank Peter Dwyer for calling the competitive exclusion principle to my attention.

[3] Thus non-transitivity is a problem for the perceptual model as well as for the taxonomic.

[4] I have described this extension of the perceptual model as either formally precise or reasonably interpretable but not both. If we define distances as real numbers and the classification space as Euclidean, the idea of ratios between distances is formally precise. However, what is the meaning of a distance of 38.47 in classification space? On the other hand, an imprecise natural language sense of the notion of distance is a powerful analogy which may prove useful for understanding the processes of folk biological classification.

[5] It would also be convenient for $CR < O$ when G^x_{min} is less than an internal distance called the *maximin* (G^n_{mxn}). This is the largest (max) of the smallest (min) distances between any given point in the cluster and all other points in that cluster (see Figure 5B). This is intuitively reasonable. This limit corresponds to the criterion for a pretaxon of the first model but is less strict. Some such limiting condition can account for the existence of generic taxa which belong to no life form, for empty regions of the classification space, and for residual taxa. We may also define a "chain," i.e., a minimally distinct cluster, as a set of percepts for which $G^n_{mxn} < G^x_{min} < H$.

references cited

Berlin, Brent
    1972 Speculations on the Growth of Ethnobotanical Nomenclature. Language and Society 1:51-86.
    1973 Folk Systematics in Relation to Biological Classification and Nomenclature. Annual Review of Ecology and Systematics 4:259-271.
Berlin, Brent, Dennis E. Breedlove, and Peter H. Raven
    1968 Covert Categories and Folk Taxonomies. American Anthropologist 70:290-299.
    1973 General Principles of Classification and Nomenclature in Folk Biology. American Anthropologist 75:214-242.
    1974 Principles of Tzeltal Plant Classification: An Introduction to the Botanical Ethnography of

a Mayan-Speaking People of Highland Chiapas. New York and London: Academic Press.

Black, Mary
1967 An Ethnoscience Investigation of Ojibwa Ontology and World View. Unpublished Ph.D. dissertation, Stanford University.

Bright, Jane O., and William Bright
1965 Semantic Structures in Northwestern California and the Sapir-Whorf Hypothesis. *In* Formal Semantic Analysis. E. A. Hammel, Ed. American Anthropologist 67(5, pt. 2):249-258.

Brown, Cecil
1974 Unique Beginners and Covert Categories in Folk Biological Taxonomies. American Anthropologist 76:325-327.

Bruner, Jerome S., J. J. Goodnow, and G. A. Austin
1956 A Study of Thinking. New York: Wiley.

Bulmer, Ralph N. H.
1970 Which Came First, the Chicken or the Egghead? *In* Échanges et Communications, Mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème anniversaire. J. Pouillon and P. Maranda, Eds. The Hague: Mouton. pp. 1069-1091.

Dennler, Jorge G.
1939 Los Nombres Indígenas en Guaraní de los Mamíferos de la Argentina y Países Limítrofes y su Importancia para la Sistemática. Physis 16:225-244.

Diamond, Jared M.
1966 Zoological Classification System of a Primitive People. Science 151:1102-1104.

Durkheim, Emile, and Marcel Mauss
1963 Primitive Classification. Translated and edited by Rodney Needham. Chicago: University of Chicago Press. (Original: 1903).

Gregg, John R.
1954 The Language of Taxonomy, an Application of Symbolic Logic to the Study of Classificatory Systems. New York: Columbia University Press.
1967 Finite Linnaean Structures. Bulletin of Mathematical Biophysics 29:191-206.

Hays, Terence E.
n.d. A Method for Deriving Covert Categories with Nonliterate Informants. Paper presented at the 72nd American Anthropological Association Annual Meeting, New Orleans. 1973.

Hunn, Eugene
1973 Tzeltal Folk Zoology: The Classification of Discontinuities in Nature. Unpublished Ph.D. dissertation. University of California, Berkeley.
1975a A Measure of the Degree of Correspondence of Folk to Scientific Biological Classification. American Ethnologist 2:309-327.
1975b Cognitive Processes in Folk Ornithology: The Identification of Gulls. Working Paper No. 42. Language-Behavior Research Laboratory, University of California, Berkeley.

Irving, Laurence
1953 The Naming of Birds by Nunamiut Eskimo. Arctic 6:35-43.

Kay, Paul
1971 Taxonomy and Semantic Contrast. Language 47:866-887.
1973 A Model-Theoretic Approach to Folk Taxonomy. Social Science Information 14:151-166.

Mayr, E., E. G. Linsley, and R. L. Usinger
1953 Methods and Principles of Systematic Zoology. New York: McGraw-Hill.

Perchonock, Norma, and Oswald Werner
1969 Navaho Systems of Classification: Some Implications for Ethnoscience. Ethnology 8:229-242.

Randall, Robert
1976 How Tall Is a Taxonomic Tree? Some Evidence for Dwarfism. American Ethnologist 3:543-553.

Simpson, George Gaylord
1961 Principles of Animal Taxonomy. New York: Columbia University Press.

Turner, Nancy Jean
1974 Plant Taxonomic Systems and Ethnobotany of Three Contemporary Indian Groups of the Pacific Northwest (Haida, Bella Coola, and Lillooet). Syesis 7, Supplement 1.

LINKED CITATIONS
- Page 1 of 2 -

References Cited

Folk Systematics in Relation to Biological Classification and Nomenclature
Brent Berlin
Annual Review of Ecology and Systematics, Vol. 4. (1973), pp. 259-271.
Stable URL:
http://links.jstor.org/sici?sici=0066-4162%281973%294%3C259%3AFSIRTB%3E2.0.CO%3B2-T

Covert Categories and Folk Taxonomies
Brent Berlin; Dennis E. Breedlove; Peter H. Raven
American Anthropologist, New Series, Vol. 70, No. 2. (Apr., 1968), pp. 290-299.
Stable URL:
http://links.jstor.org/sici?sici=0002-7294%28196804%292%3A70%3A2%3C290%3ACCAFT%3E2.0.CO%3B2-%23

General Principles of Classification and Nomenclature in Folk Biology
Brent Berlin; Dennis E. Breedlove; Peter H. Raven
American Anthropologist, New Series, Vol. 75, No. 1. (Feb., 1973), pp. 214-242.
Stable URL:
http://links.jstor.org/sici?sici=0002-7294%28197302%292%3A75%3A1%3C214%3AGPOCAN%3E2.0.CO%3B2-I

Semantic Structures in Northwestern California and the Sapir-Whorf Hypothesis
Jane O. Bright; William Bright
American Anthropologist, New Series, Vol. 67, No. 5, Part 2: Formal Semantic Analysis. (Oct., 1965), pp. 249-258.
Stable URL:
http://links.jstor.org/sici?sici=0002-7294%28196510%292%3A67%3A5%3C249%3ASSINCA%3E2.0.CO%3B2-R

LINKED CITATIONS

- Page 2 of 2 -

Unique Beginners and Covert Categories in Folk Biological Taxonomies
Cecil H. Brown
American Anthropologist, New Series, Vol. 76, No. 2. (Jun., 1974), pp. 325-327.
Stable URL:
http://links.jstor.org/sici?sici=0002-7294%28197406%292%3A76%3A2%3C325%3AUBACCI%3E2.0.CO%3B2-E

Zoological Classification System of a Primitive People
Jared M. Diamond
Science, New Series, Vol. 151, No. 3714. (Mar. 4, 1966), pp. 1102-1104.
Stable URL:
http://links.jstor.org/sici?sici=0036-8075%2819660304%293%3A151%3A3714%3C1102%3AZCSOAP%3E2.0.CO%3B2-W

Taxonomy and Semantic Contrast
Paul Kay
Language, Vol. 47, No. 4. (Dec., 1971), pp. 866-887.
Stable URL:
http://links.jstor.org/sici?sici=0097-8507%28197112%2947%3A4%3C866%3ATASC%3E2.0.CO%3B2-%23

How Tall Is a Taxonomic Tree? Some Evidence for Dwarfism
Robert A. Randall
American Ethnologist, Vol. 3, No. 3, Folk Biology. (Aug., 1976), pp. 543-553.
Stable URL:
http://links.jstor.org/sici?sici=0094-0496%28197608%293%3A3%3C543%3AHTIATT%3E2.0.CO%3B2-R