how tall is a taxonomic tree? some evidence for dwarfism[1]

ROBERT A. RANDALL—*University of Victoria*

Descriptions of human belief are usually psychologically inaccurate because they fail to distinguish mentally stored old knowledge from new knowledge spontaneously produced by reasoning. The result of "lumping" such disparate psychological phenomena is often a tendency to describe as encyclopedic what could more accurately and economically be derived from fundamental premises and reasoning rules. Although such oversimplification is common in ethnographic analysis, here I will only consider its implications for the description of folk biological classification.

In most such studies, we are presented with a taxonomic "tree" diagram (see Figure 1). Starting from each lowest terminal taxon, there is a chain of more inclusive taxa proceeding upwards through other taxa to a "unique beginner." Thus, as in Figure 1, a *black oak is a kind of oak, an oak is a kind of tree,* and *a tree is a kind of plant.* Alternatively, the diagram may be interpreted in a different fashion: starting from a unique beginner, there is a chain of "direct precedence" proceeding downward through other taxa to terminal taxa. Thus, the taxon, *plant,* includes the taxon, *tree, tree* includes *oak,* and *oak* includes the taxon, *black oak.*

To the extent that these diagrams reflect informant statements of the sort "X is a kind of Y," there can of course be no quarrel. But there is a covert aspect of these diagrams which goes considerably beyond a mere summary of informant responses. Specifically, the relationship between *black oak* and *tree, black oak* and *plant,* and *oak* and *plant* need not be modeled indirectly. If we are only trying to describe the fact that informants claim that *black oaks are trees and also plants,* and that *oaks are plants,* we can, as in Figure 2, more accurately place *black oak* directly under both *tree* and *plant* and place *oak* directly under *plant.* Thus, if the indirect precedence described in Figure 1 is to be considered

Several arguments are made in this paper: (1) Taxonomic tree models of folk classification are implicitly generative because they produce appropriate statements which are not in the description itself. (2) The generative devices sometimes postulated—namely, transitive reasoning operating on chains of directly included taxa—do not account for some evidence which another model, the direct comparison between prototypic images, does. (3) Taxonomic trees are probably not stored directly in the memory except perhaps as "dwarf" trees consisting of contrast sets and their names. (4) Routine classification behavior is not so much a matter of producing giant taxonomic trees as it is a matter of selecting, in particular socioeconomic situations, a characteristic of an organism relevant for action.

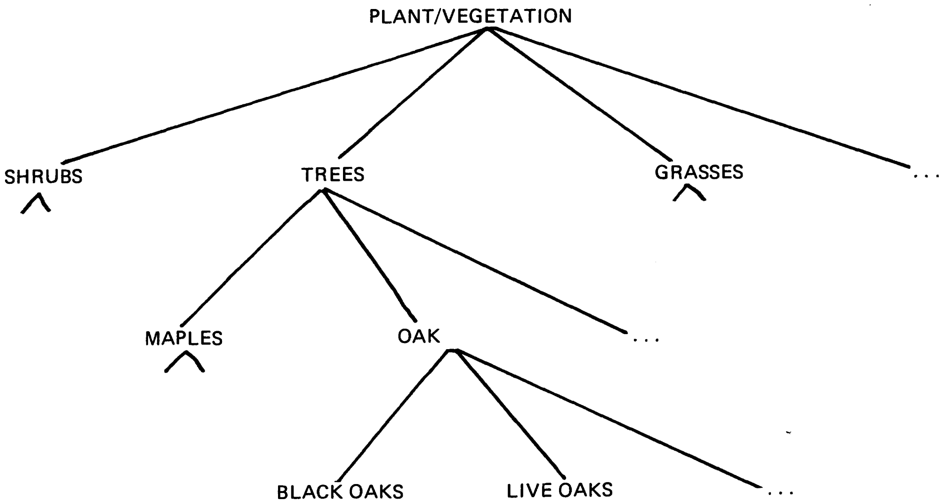

Figure 1. Example of standard folk taxonomic tree.

more accurate than the spider web of Figure 2, it is necessary to make some additional cognitive claim for the model of Figure 1.

Although to my knowledge, this issue has not previously been explicitly raised, it seems safe to assume that most of us would argue as follows:

Figure 2 implies that the fact that *a black oak is a kind of tree* is stored directly in the memory. This seems unlikely because direct storage would burden the memory unnecessarily. Moreover, interview data does not support such a model of information storage. When informants are asked to name all the kinds of *trees* they can think of, they do not usually respond with a general taxon, then a specific one, then a more specific one, then a general one, then a more specific one, then another general one, and so forth. Thus informants do not as a rule answer, "Well, there are oaks, black oaks, pines, white pines, live oaks, evergreens, ... " Rather, they just give "generic rank taxa": *oaks, maples, pines*, and so forth. For this reason, we can be sure that *black oaks* are stored in the memory under *oaks*, not *trees* or *plants*.

But if *black oaks* are not stored in the memory directly under *tree*, there must be some sort of reasoning process that tells us that *black oaks are trees*. In this respect, taxonomic descriptions are implicitly generative because they produce appropriate indirect precedence statements not already stored in the memory. For the most part, however, there has been little speculation on how this reasoning takes place. Presumably, as Kay implies (1971),[2] direct precedence relations produce indirect ones by the application of transitive logic: "If a black oak is a kind of oak, and an oak is a kind of tree, then a black oak must be a kind of tree."

Despite the simplicity and obvious truth of this syllogism, there is to my knowledge absolutely no empirical support for the existence of such reasoning. In short, then, diagrams of the type in Figure 1, when interpreted to imply memory storage of direct precedence relations and the use of transitive reasoning to generate indirect ones, must be considered plausible explanations of facts, but not facts themselves.

Nevertheless, people *are* generally capable of reporting these nonmemorized inclusion relationships, so it may be useful to consider whether better alternative explanations for these abilities exist.

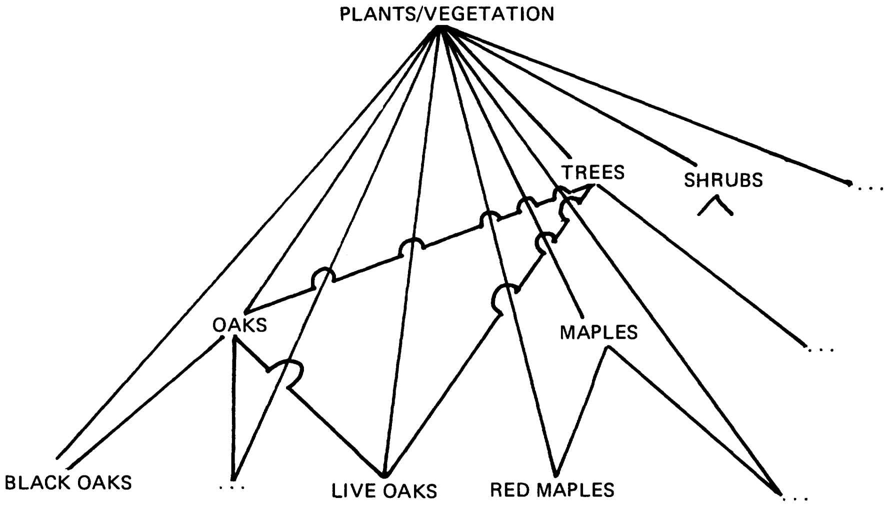

Figure 2. Direct memory storage of taxonomy.

some taxonomies are created spontaneously by informants

I first began to doubt the status of some taxonomies while engaged in studies of kin terminology. In the English language, the Philippine language of Sinama,[3] and probably other languages as well, a taxonomy of "relatives" can easily be elicited, but numerous studies have conclusively shown that kin terminology is stored in the memory in some other fashion. Thus, for example, in English, the category, *ancestors* (a kind of *blood relative*), has an indefinitely large number of immediately included subordinate taxa such as *great-great-grandmother*, etc. Obviously, these categories cannot be directly stored in a finite memory, so they must be generated by some productive process. In general, then, taxonomies of "relatives" are most likely created on the spot by the productive use of feature definitions, extension rules, and directly stored genealogical categories (cf. Lounsbury 1964).

In a subsequent economic and ecological study among the Samal, I found it comparatively easy to elicit marine food chains from fishermen. All I had to do was take a sufficiently large marine organism, such as *kaitan* ('shark'),[4] and ask "What are all the things sharks eat?" Then, with ready answers such as *kaitan nahut* ('small sharks'), *panit* ('tuna'), *maŋkoʔ* ('frigate mackerel'), and the like, I was able to apply the 'What-are-all-the-things-X-eat?' frame to the 'sharks' ' prey. The questioning process was then continued recursively until answers naming very small forms of marine life were given. Then, with such "primary consumers" as *daiŋ nahut toʔod* ('very small fish'), "producers" such as *lumut batu* ('rock scum-algae'), and the "abiotic substance" *boheʔ* ('water'), my informants reached the terminal consumable of the chain.

Although I elicited knowledge that was clearly stored in the fishermen's memories, I do not think it correct to say that it was stored *directly* (cf. Miller, Galanter, and Pribram 1960:169-180). My literate informants were much too surprised at my diagram for this to be old knowledge. Certainly, they "knew" all these facts about "sea life," but I doubt if they had ever put them together in this way before. It had probably never occurred to them before that ultimately 'sharks' depend for their food supply on the consumption of 'water' by 'rock scum-algae.'

It is probably a similar set of circumstances which necessitates that our own society's ecologists teach that all life depends on the fixation of solar energy. The fact is easily demonstrated by transitive reasoning, but for many individuals in our society, there is simply never any situation that warrants the application of such logic. Hence, ecologists must constantly try to show people that ultimately their lives depend on plants and solar energy. However, unlike ecologists, ordinary people have little or no interest in food chains, so they do not bother to store them directly.[5] Probably, like taxonomies of relatives, the folk food chain is an artifact of an irrelevant questioning procedure. It is most likely a true but culturally unimportant statement of previously unsuspected relationships.

I do not remember the science class where I first learned about food chains, but I do remember in my early anthropological days being surprised to discover that I had a taxonomy of *something to eat* in my head. Subsequently, I was startled to realize that all sorts of weird taxonomies could, like a tooth, be pulled from my lips: *kinds of relatives, kinds of plants, kinds of furniture, kinds of weapons,* and even *kinds of university employees.* I was surprised that I knew these things I think, not because the taxonomies violated my intuitions, but because I had not previously realized that I knew these things so systematically.

Of course, there is a ready explanation for this sense of surprise. Labeled segregates are stored taxonomically in the memory by unconscious processes. Hence, we are surprised when an ethnographer makes what was previously unconscious, conscious. However, there is also another explanation—that folk taxonomies as we know them from ethnographic tradition are, like folk food chains, constructed by applying normally unexploited principles of native logic to various scraps of knowledge lying around in the mind. In this view, then, a folk taxonomy would be more like a previously undiscovered geometry theorem than a basic axiom of geometry.

do multi-level taxonomies exist in the memory?

As we shall see, there is some evidence that taxonomies are artifacts of more basic knowledge. For one thing, in both English and Samal there are folk biological classifications in which indirect inclusion relations are not transitive. Thus, an American friend of mine routinely lists *oaks* as kinds of *trees.* Nevertheless, while hiking with me in the mountains of California, he had occasion to tell me that *scrub oaks* are not *trees* but *shrubs.* I realized that the facts were nontransitive, but let them pass because diversified families are common in botany. Subsequently, however, another informant gave *berries* as an example when asked "What are the kinds of bushes?" But when I later asked, "What kinds of berries are there?" he listed *raspberries, blackberries, strawberries, blueberries,* and several others. Needless to say, he does not believe that "strawberries are kinds of bushes." In the course of ascertaining this, I pointed out to him that it is illogical to say in effect: "Strawberries are kinds of berries. Berries are kinds of bushes. But strawberries are not kinds of bushes." In response, he just laughed and said, "Berries are *generally* bushes." Apparently, this gross lack of transitivity hardly concerned him at all!

Indeed, similar logical quandaries do not seem to bother me either. *Beans* are *vines,* and *bush beans* are kinds of *beans,* but *bush beans* are most certainly not *vines; willows* are *trees,* and *pussy willows* are *willows,* but many *pussy willows* in the Western United States are not *trees,* but *shrubs.*

Among the Samal I worked with, all 'non-sentient life' is categorized dichotomously as either *isi gumi* ('flesh of the land') or *lumput laut* ('sea flesh'). Figure 3 shows that one type of land flesh is *tumbutumbuhan* ('vegetation') and that *sagbot* ('nonwoody vegetation') is a kind of vegetation. 'Agar-agar,' 'sargassum,' and other *sagbot tahik* ('seaweeds') are included within the 'nonwoody vegetation' category. However, 'seaweed' is not considered to be 'flesh of the land.' Thus, seaweed is a kind of nonwoody vegetation, and nonwoody vegetation is a kind of land flesh, but seaweed is not a kind of land flesh.

Similarly, in Figure 3, the label *lumut,* meaning literally 'scum,' but mainly including various lichens and other minute green organisms, is classified as 'flesh of the land,' probably because it is regularly seen on rocks and coconut trunks. However, an occupationally important minute sea algae is a kind of scum, scum is a kind of land flesh, but minute sea algae is not a kind of land flesh.

In fact, though, standard folk biological theory gives little hint as to how such nontransitive thoughts might be stored in memory. One might say correctly that 'land flesh' is not a configurational category and thereby argue that it is outside the taxonomic

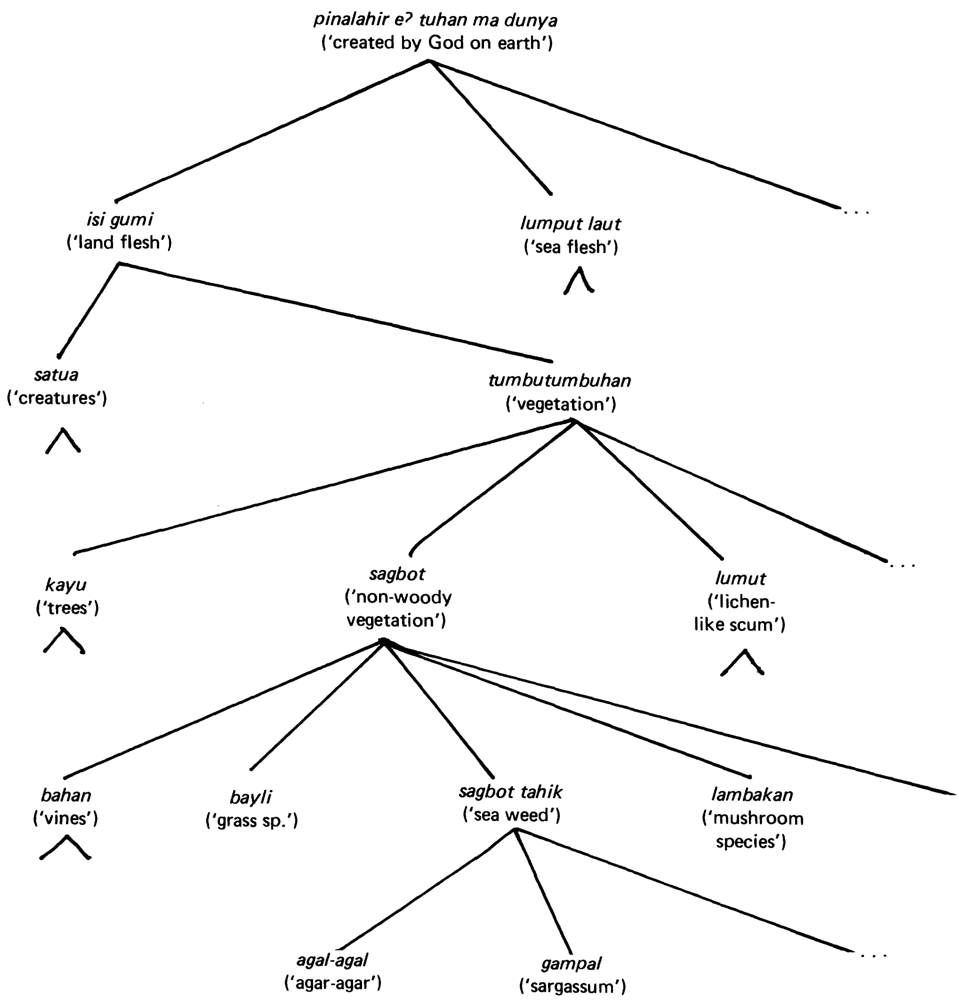

Figure 3. A partial Samal taxonomic tree
of things created by God on earth.

classification schema. As support for this position, it could be argued with some justification that only configurational categories should be transitive because only they reflect the evolutionary reality of sense data. Nontransitive configurational triples like *scrub oak, oak,* and *tree* could be dismissed by arguing that they are rare in communities where little intentional genetic modification of species has occurred, where the importation of organisms has not been fostered, and where individuals seldom travel into different ecosystems.

Nevertheless, it could also be claimed that nontransitives are evidence that people do not routinely use transitive reasoning at all. If this were the case, one could argue that nonconfigurational categories are systematically related to the folk classification system, and that therefore they cannot be arbitrarily excluded just because they are less likely to be transitive. In this view, then, it is the multilevel taxonomy model that should be discarded, not the nonconfigurational facts.

Since people must have some means of knowing that *black oaks* are *trees* and that *scrub oaks* are not, an explanation of such knowledge is needed. If the standard taxonomic hypothesis is accepted, the options are limited. One could suppose that two precedence chains such as *scrub oak* ⟵ *oak* ⟵ *shrub* and *black oak* ⟵ *oak* ⟵ *tree* are stored as "crosscutting" hierarchies as in Figure 4. But clearly, this is unacceptable

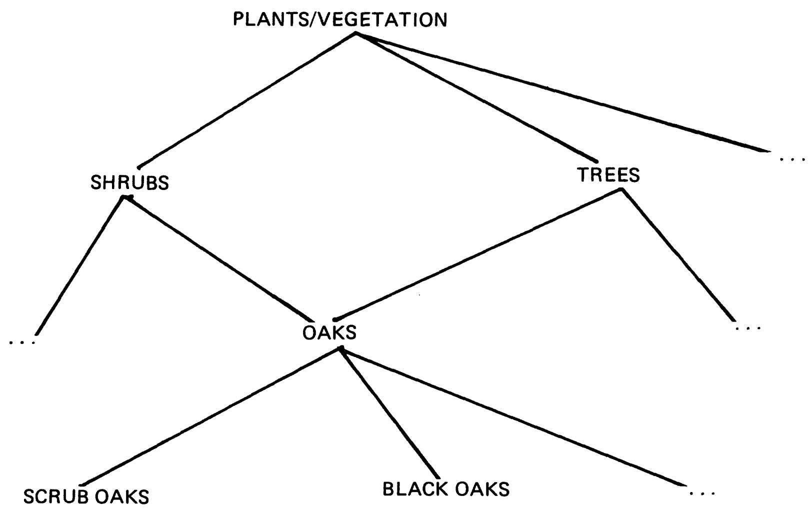

Figure 4. A crosscutting hierarchical model.

because with this type of storage, it would be possible for an individual to think "scrub oaks are trees." Obviously, then, if standard transitive logic is to be preserved, *oak* must be stored as two separate categories (see Figure 5). In this way, an individual could use the reasoning to determine unambiguously the life-form status of *scrub oaks* and *black oaks*. Thus, if we accept the standard transitivity hypothesis, we also must accept Figure 5 and its concomitant assumption that *oak* is stored in two different places.

Of course, as Kay (n.d.) suggests, it may be possible to modify the transitivity hypothesis without completely discarding it. One could argue, as in Figure 6, that only "typical kinds" are to be considered to be taxonomic categories. However, this approach still does not explain how we generate indirect statements about typical taxa, and it

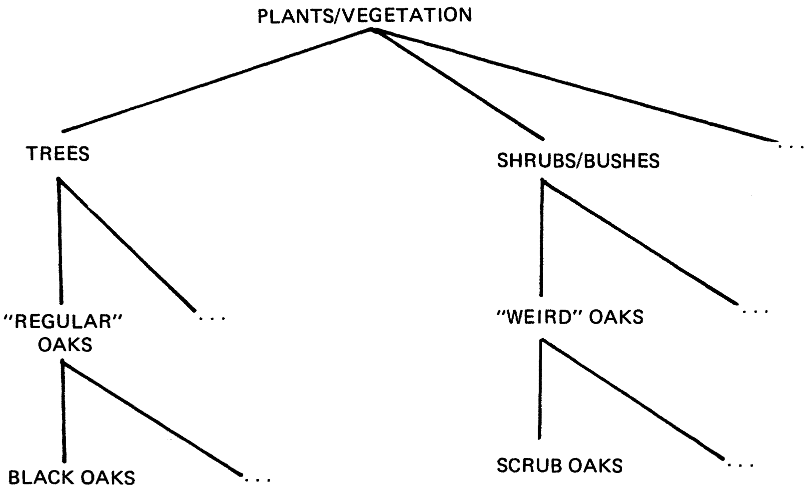

Figure 5. A separate memory location model.

ignores the classification of atypical taxa entirely. Thus, for example, Figure 6 does not tell us that *scrub oaks* are *oaks*.

There may be other more complicated ways to modify the transitivity hypothesis and still account for these particular verbal facts. However, in my opinion, an intuitively more appealing alternative is to suppose the following: non-scholastically trained people seldom, if ever, store large taxonomic trees in memory directly. Rather, when necessary, they can recall the perceptual characteristics of classes of living organisms and use this knowledge purposefully for gathering acorns, landscaping a garden, naming plants, creating classification schema, and so forth.

Classification could be accomplished by the recall of perceptual characteristics in at least two ways. If we assume that certain individuals store configurational images of "the typical scrub oak," "the typical tree," and "the typical bush" directly in the memory, then it seems possible that some sort of direct comparision between these images could be

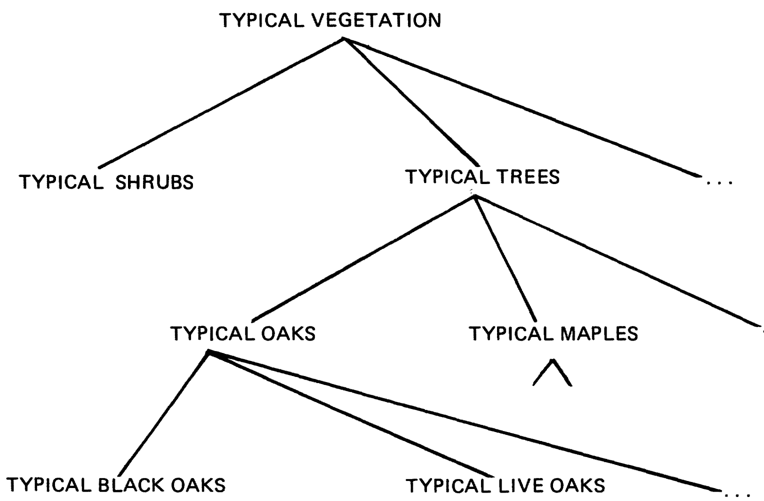

Figure 6. A "typical kinds" taxonomic tree model.

made. Thus, in order to determine whether a *scrub oak* is *tree-like,* the two images could simply be recalled from the memory and their shapes compared. Presumably, a similar ability to compare images is used in selecting building materials, so it is at least plausible that the same ability is sometimes invoked when people classify.

On the other hand, it seems possible that such comparisons are often unnecessary. Perhaps associations between categories and their perceptual characteristics can be stored directly in the memory. Thus, for example, the category *scrub oak* might have associated with it in the memory characteristics which we could call "dwarfness," "oakness," and "shrubness." A *weeping willow* category would have associated with it the drooping, dangling aspect that we also find in *weeping cherries* and *weeping beeches.* A *black oak* might be said to have not only the life form and generic characteristics, "tree-like" and "oakness," but also a configurational quality that might be labeled "oak-blackness' (to contrast with other leaf-and-bark configurations such as "maple-blackness" and "willow-blackness").

If the memory contained such associations between plant categories and their configurational characteristics (see Figure 7), then it would be comparatively easy for an informant to answer the usual type of hierarchically framed questions, even though the knowledge is not stored in this way. If, for example, an informant were asked to name *kinds of bushes,* he would merely search his memory for generic and specific characteristics associated with bushiness. Since most *berries* are bushy, he would probably mention them near the beginning of the list; since most *oaks* are not bushy, he would probably not mention them at all. On the other hand, if asked whether all *berries* are *bushes,* the informant would need a different strategy. Instead of searching *bushes* for their generic characteristics, he would be searching *berries* for their life form characteristics. Presumably, such a search would easily determine that *strawberries* are not *bushes.* In effect, then, both typical and atypical flora can be classified with the memory association model.

Such a model also explains other data. If non-configurational characteristics (such as "use" or "ecozone location") were also associated with categories in the memory, then certain types of extremely common classification could be explained. The ability to list (or choose among) the kinds of *vegetables,* the kinds of *wild flowers,* or the 'kinds of sea flesh' clearly depend on the storage of essentially nonconfigurational characteristics. Moreover, rather ordinary ecological generalizations depend on the same type of memory storage: "Oaks are typically trees at low elevations but are typically bushes at higher elevations"; "Willows are typically bushy along Western streams but are typically trees in Eastern gardens." Although it could be argued that oaks are typically trees in "typical" ecosystems and that willows are typically bushes in "typical" ecosystems, such assertions seem forced. Ordinary people are well aware that typicality is sensitive to many types of information in addition to configuration.

In any given case, then, an individual might use either directly stored associations of characteristics or a comparison of prototypic images to perform a classification task. Probably most people determine the "tree" status of pussy willows by comparing images. This is because they have never had to solve this problem before. On the other hand, providing the "oak" status of *black oaks* is probably easier, because, as the label suggests, the "oak-blackness" characteristic is directly associated in memory with "oakness." *Black oaks* may be classified merely by recalling the association.

In short, then, both the comparison of images hypothesis and the association of characteristics hypothesis explain more classification behavior than does the taxonomic hypothesis with its empirically unmotivated assumption of transitive reasoning. Thus,

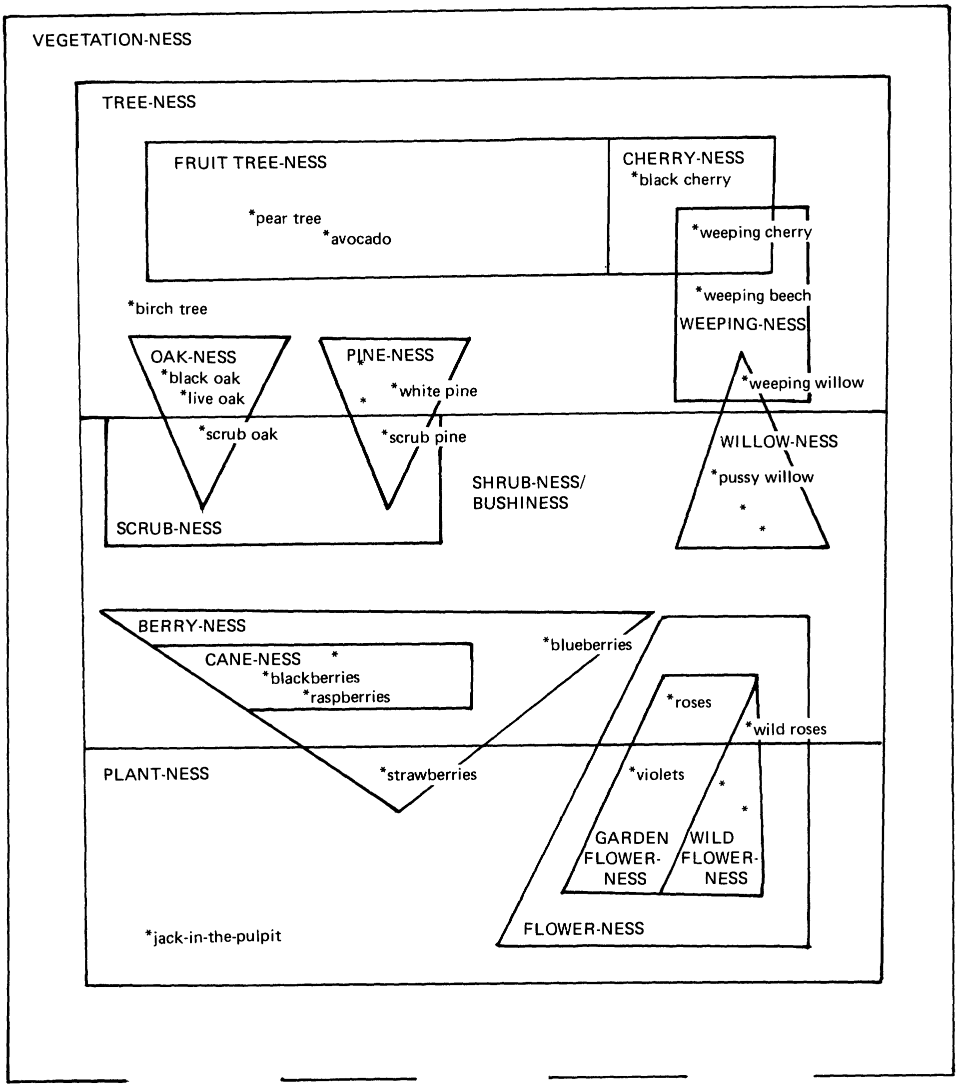

Figure 7. A memorization of characteristics model of some English plant categories.

they better account for what people are concerned with (the important characteristics of living things) and what people are unconcerned with (the nontransitive relations existing between some categories). More importantly, they provide alternatives to the taxonomic tree hypothesis. Therefore, it is no longer justifiable to *assume* that taxonomies are directly stored in the memory, except, perhaps, in part as "dwarf" trees (*oak* and *black oak*) having a direct precedence length of two.

Lack of evidence does not prove a hypothesis wrong, however. It therefore seems useful to determine whether hard data can be found to support one model or undermine another. To me, the taxonomic tree hypothesis seems wrong because introspection tells me that I do not use transitive reasoning to determine, for example, that "blue grass is a kind of vegetation." Rather, I know that *vegetation* has certain configurational properties that *blue grass* also shares. Such intuition is supported by analogous field data. In particular, I have often found that informants compared characteristics verbally when asked to provide a particular classification. For example, I have in my Samal field notes a justification of the belief that *lambakan* ('an edible mushroom'), a form of *sagbot* ('nonwoody vegetation'), is a kind of *tumbutumbuhan* ('vegetation') (see Figure 3). The informant did not claim that it was *sagbot* and then use this fact to argue that it must be vegetation. Rather, he merely pointed out that mushrooms, like other vegetation, have what he considers to be 'stems,' 'roots,' and 'leaves.' In effect, he claimed that mushrooms can be considered vegetation because they have a vegetative image.

This is not the only evidence in support of my model. For example, when asked to name *kinds of berries,* my American informant included *strawberries* with *blueberries* and *blackberries.* He did not forget to mention *strawberries,* nor did he name them only after exhaustively listing all *berries* that are *bushes,* as he might have done if he stored *strawberries* in a memory location distinct from *typical berries.* Obviously, his behavior would be anticipated if the association of characteristics in Figure 7 were stored, but it would be unlikely if the separate chains of Figure 5 were in his memory.

More generally, when we ask people to name kinds of things, they usually respond by listing contrasts in one type of characteristic. Thus, when asked to list kinds of vegetation, Americans and Samals tend to list life-form characteristics (trees, mushrooms, etc.), use characteristics (vegetables, weeds, etc.), habitat characteristics (wild flowers, garden flowers, etc.), or generic characteristics (tomatoes, beans, etc.). By contrast, as far as I know, they never make their way systematically through the nodes of a taxonomic tree:

> Well, there are trees. Evergreens are kinds of trees. Pines are one sort of evergreen. Lodgepole pines are kinds of pine.

In effect, then, memory recall data suggest that vegetation categories are stored in the memory primarily as features (prototypic or otherwise) in contrast sets.

In my opinion, this finding points toward a solution to the problem posed by the taxonomic tree hypothesis. If storage is by contrast sets, then taxononomic trees can only be the result of classification *behavior.* Obviously, any characteristic can be further subdivided into finer characteristics, so any competent person who wanted to should be able to create hierarchies of characteristics. Triples of features such as (color, red, scarlet), (relative age, my generation, elder than me), and (class, upper class, millionaire) can be provided by any English-speaking actor competent enough to make broad distinctions in some situations and narrow distinctions in others.

But what motive would an individual have for doing this? This, I think, is the crux of the problem. Instead of consciously systematizing, most people tackle a different task. It seems to me that the important classification problem routinely facing intelligent humans is to operate adequately in a physically demanding, complex, and often dangerous socioecological environment. Doing this does not involve constructing taxonomic trees, but rather, in a particular situation, selecting a contrast set of characteristics which is both sufficiently specific to achieve a practical and safe result and sufficiently general to accomplish one's purposes efficiently. If this is the primary use of classification, then it is in the mechanisms of choice and in the historical context of the choice that we must look for an explanation of classification.

conclusion

Although it may be possible to propose some transitive-like rule which describes taxonomic naming, there seems little motivation for doing so. In fact, there are better ways to explain memory storage and natural classification behavior. Of course, "giant" taxonomic trees are "real" in the sense that they can be produced by people (and certainly, this is fortunate for those of us trying to discover categorical domains during fieldwork). But how and why are such trees spontaneously produced? Given the data cited above, it seems likely that they are not fashioned from transitive logic and they are not very often used in everyday life. Presumably, then, those who still believe that giants dominate the classification forest should be prepared to produce *evidence* that this is so. Otherwise, the suspicion will linger that believers in giants are simply barking up the wrong tree.

notes

[1]Although the position taken here is entirely my own, I have also benefited from discussion on these topics with several colleagues: Brent Berlin, William Geoghegan, Hugh Gladwin, Gene Hunn, Paul Kay, and Nancy Edwards-Randall. Amos Tversky (1969) helped convince me that transitivity assumptions are not always warranted.

[2]The lefthanded implication of Kay's (1971) theorem 15 reads as follows: if t_i occurs in the sequence $t_1 \longrightarrow \ldots \longrightarrow t_j$, then for distinct t_i, $t_j \epsilon T$, $t_i \supset t_j$. [Here t_i and t_j are taxa in the taxonomic structure T, and $\longrightarrow$ denotes direct precedence.] If one accepts this theorem as an attempt to model the reasoning process (and Kay makes no such claim), then we can only assume that an individual uses knowledge of the direct precedence chain in order to determine whether or not t_i includes t_j. In the proof of this theorem, Kay shows that this competence depends on the transitive property implied in the formal notion of inclusion.

[3]Fieldwork among the Samal of Linuŋan Island, Basilan Strait, Southwestern Mindanao, was conducted from August 1971 to September 1972. I wish to acknowledge the support of the inhabitants of that island, as well as those who funded the project: the U.S. National Science Foundation and the U.C. Berkeley Institute of International Studies. In this paper, all references to the terms "Samal" and "Sinama language" should be understood to refer solely to the Linuŋan situation.

[4]The systematic icthyological referents corresponding to the 'fish' named in this paper are as follows: *kaitan* ('shark,' i.e., longish members of the Class Elasmobranchii such as the orders Lamiformes and Rhinobatiformes); *panit* ('tuna,' i.e., Family Thunnidae); and *maŋkoʔ* ('frigate mackerel,' i.e., Auxis thazard).

[5]That is, they do not commonly store specific food chains in memory. As a reviewer of this paper points out, there may be some sort of general recognition that big fish eat smaller ones. Thus, for example, a Samal barrio official told me that an employee of the Philippine Department of Fisheries had explained to him that "explosives" fishing is bad because it kills the small fish that larger fish use for food. However, other Samal fishermen seemed either unaware of this fact or at least uninterested in it. They uniformly ascribed low catches in dynamited areas entirely to the fishes' fear of explosives.

A more common case of food chain reasoning revolves around the eating habits of ducks. Several Samals told me that they would not eat duck flesh because "ducks eat human feces." As far as I can tell, their notion of Islam requires them to consider the food chains of the organisms they eat. I doubt if such food chains are ever extended more than two trophic levels.

references cited

Kay, Paul
1971 On Taxonomy and Semantic Contrast. Language 47:866-887.
n.d. A Model-Theoretic Approach to Taxonomy Theory. A paper presented at the 72nd Annual Meeting of the American Anthropological Association, New Orleans, 1973.
Lounsbury, Floyd
1964 A Formal Account of the Crow- and Omaha-type Kinship Terminologies. *In* Explorations in Cultural Anthropology. Ward Goodenough, Ed. New York: McGraw-Hill. pp. 351-393.
Miller, G. A., Galanter, E., and K. Pribram
1960 Plans and the Structure of Behavior. New York: Holt, Rinehart, and Winston.
Tversky, Amos
1969 Intransitivity of Preferences. Psychological Review 76:31-48.

Date of Submission: December 10, 1974
Date of Acceptance: April 1, 1975