Chapter 4

CATEGORIES AND COGNITIVE ANTHROPOLOGY*

JAMES S. BOSTER

University of Connecticut

Contents

* My understanding of these issues has been greatly influenced by Roy D'Andrade, to whom I am very grateful. I would also like to thank him, Cornelia Dayton, Paul Kay, Asifa Majid, and Kateryna Maltseva for their comments on this chapter. I am grateful to David Kenny and John Gatewood for their help in locating Campbell's writing on "entitivity." I thank Claire Lefebvre and Henri Cohen for their hard work in compiling the volume and Zofia Laubitz for her assiduous copyediting.

Handbook of Categorization in Cognitive Science, Edited by Henri Cohen and Claire Lefebvre

Abstract

In this chapter, I review how choices of methods of data collection and analysis have affected the representations of category structure by cognitive anthropologists in three domains: kinship terminologies, color classification, and ethnobiology. I argue that cognitive anthropologists have contributed to cognitive science in part by developing a "science of the stimulus," leading to a better understanding of how the nature of the objects of experience affects the ways humans categorize them. I show that this approach offers an explanation of why some forms of semantic organization of lexicons are rare (e.g., paradigms) while others are common (e.g., trees and taxonomies), in part by expanding Campbell's (1958) criteria for entitivity to include both interdependence and discreteness of features. Domains whose features are independent of each other and vary continuously lack the entitivity, in most cases, to be recognized as a class of things worth classifying. This observation, plus a comparison of the lexical overhead of naming unique combinations of feature values rather than the feature values themselves, helps explain the rarity of paradigms. Trees and taxonomies, conversely, are much more common because the domains they lexicalize have much greater entitivity by these amended criteria. I discuss other examples of important advances in the cognitive anthropological understanding of particular domains that occurred when the methods and assumptions of the analysis better matched the nature of the domain being studied. In general, a "science of the stimulus" prompts the investigator to ask, "Where does domain structure come from?" In the case of ethnobiology, both the correlational structure of the domain and the cognitive architecture that perceives and categorizes it are crafted by the process of evolution by natural selection.

1. Introduction

This chapter reviews some of the contributions cognitive anthropologists have made to the understanding of human categorization. Rather than duplicate Roy D'Andrade's (1995) overview of the history of the development of the field, I will focus on two questions. First, how do choices of methods of data collection and analysis affect the representations of category structure in various domains? Second, how do cognitive anthropologists contribute to a "science of the stimulus?"

The chapter begins by articulating the difference between cognitive anthropology and the prototype of the cognitive sciences – cognitive psychology – and presenting some historical background. This is followed by an examination of a portion of what cognitive anthropologists have learned in three domains: kinship, color, and ethnobiology[1]. Although this review leaves out much more of the field than it includes, it will be used to answer the questions posed above.

2. Cognition and culture, universalism and relativism

If cognitive psychology is the study of the *process* of thought in *individuals,* observed in *experimental* settings, cognitive anthropology is the study of the *content* of thought (or knowledge) in *communities* of individuals observed in *natural* settings[2]. Cognitive anthropologists seek to understand how culture happens, to explore how collective understandings of the world emerge in social groups, and to discover the pattern of cross-cultural similarities and differences in culture and cognition. The field sits at the intersection of psychology and anthropology, bridging studies of individual and collective representations.

The creation myth of cognitive anthropology might start, "In the beginning there was Murdoch[3]." After the Second World War, a very impressive collection of graduate students and junior faculty members gathered at Yale under the nominal dominion of George Peter Murdoch. Many of them, like Roberts, Conklin, and Goodenough, had served in the military in the war and were far more mature and toughened than the typical entering graduate students or assistant professors of other eras. Murdoch's enterprise of cross-cultural comparison was well under way, producing *Social Structure* in 1949 (comparing 250 societies) and the "World Ethnographic Sample" in 1957 (comparing 564 societies). Although Murdoch himself had an encyclopedic knowledge of world cultures and devoured ethnographies at a pace that lesser souls might expend on

[1] A discussion of cross-cultural variation in emotion categories can be found in another chapter of this volume.

[2] Of course, there are many exceptions to this generalization – such as cognitive psychologists who do fieldwork and cognitive anthropologists who perform experiments – but the contrast as stated captures the difference in the central tendencies of the two disciplines.

[3] I have heard various versions of this creation myth from Brent Berlin, Hal Conklin, Roy D'Andrade, and Jack Roberts. Whatever errors there are in this retelling are my own.

trash novels, he left the drudge work of coding the societies for the purposes of the ethnographic atlas largely to his graduate students. Jack Roberts recalled sitting in the library beside a tall stack of ethnographies, laboriously deciding what kind of wall construction was used for the houses, what kind of marriage and kinship system the society had, etc. The talent and intelligence of the graduate student workers combined with the immensity and drudgery of the task (and perhaps some elements of Murdoch's personality) both grounded the students in a deep understanding of vast chunks of ethnographic literature and created the circumstances ripe for rebellion.

The rebellion came in the form of a sharp critique of the categories used to code the societies. The graduate students (by 2005, they are mainly distinguished professors emeriti) felt they were often forced to assign codes to societies in ways that did little justice to how the members of the society understood themselves. They wanted to develop methods that would allow the native's point of view to emerge intact, without subjecting the description of a society to the procrustean violence of artificial coding categories. The rebellion produced something that would seem very strange in later decades: a fierce devotion to the possibility of extreme cultural relativism coupled with as fierce a faith in the power of the right formal methods of systematic data collection and analysis to deliver the native models intact. Thus, even in their rebellion, these early ethnoscientists bore the imprint of their association with Murdoch. At the same time that they rejected many of the assumptions that underlay Murdoch's comparative enterprise, they embraced the rigor that characterized his systematic approach to cultural materials.

A further irony was that the discipline they started, which initially had been so committed to the possibility of extreme cultural relativism, found not only cultural differences but cultural universals as well. The tension between universalism (the notion that humans in different social groups, regardless of their natural, social, or linguistic environments, understand the world in essentially similar ways) and relativism (the notion that humans in different social and linguistic groups understand the world in very different, even incommensurate, ways) has created a great deal of debate in anthropology and, in the process, has probably wasted a great deal of ink and paper. Obscured by the debate is the fact that documentation of either position requires careful systematic data collection, analysis, and comparison. In developing the tools that could recover the native models intact in their distinctness, the early cognitive anthropologists also created the means to show that some native models were not limited to any one particular group of natives.

3. Paradigms and taxonomies

If one were to assign a date to the birth of cognitive anthropology, it would be the day Ward Goodenough delivered the paper that contained his classic definition of culture [Goodenough 1957 (1964, p. 36)]:

> A society's culture consists of whatever it is one has to know or believe in order to operate in a manner acceptable to its members, and to do so in any role that they accept for any one of themselves. Culture, being what people have to learn as distinct from their biological heritage, must consist of the

> end product of learning: knowledge, in a most general, if relative, sense of the term. By this definition, we should note that culture is not a material phenomenon; it does not consist of things, people, behavior, or emotions. It is rather the organization of these things. It is the forms of things that people have in mind, their models for perceiving, relating, and otherwise interpreting them.

Some aspects of this definition were fairly standard. Most anthropologists would have agreed that culture includes knowledge and belief and that culture is learned. However, Goodenough's insistence that both material artifacts and behavior in and of themselves were not cultural was unusual, and his assertion that culture consists of the models people use to understand the world and to guide their own behavior was novel. Also novel (for anthropologists) was the fact that the definition was an operational one. It outlined what a cultural description should look like and described a test for the adequacy of that description. An ethnography, according to Goodenough, should present a theory of the conceptual models people use rather than just a simple description of their behavior. It should be like a grammar or rule book for acting like a native. One tested the adequacy of an ethnography in much the same way that Turing thought we should test for artificial intelligence. If using the ethnography allows us to pass as a native, it is good enough[4].

Goodenough's definition gave a very wide field for exploration – "whatever it is one has to know or believe ..." – but the early cognitive anthropologists[5] generally limited

[4] One of the impacts of Goodenough's definition was to show that how one defined culture could be a tool for articulating what one thought was important about human understandings of the world and how one thought it should be studied. He joined a minor cottage industry among anthropologists, each of whom had been defining culture in their own way. However, I interpret his contribution as saying, in effect, not that culture *must* be understood in his way, but rather that if one uses his definition, one will find out about a certain kind of aspect of how humans understand the world – the model-constructing, rule-seeking, grammar-building part. His contribution served as an invitation to other anthropologists to define culture in any way they find useful, as long as it bears enough resemblance to other definitions of culture to ensure that one is talking about roughly the same sort of thing. Definitions of culture by later cognitive anthropologists have tended to bear a family resemblance to Goodenough's and to share with Goodenough's the property of articulating what they think a cultural description should be about. For example, here is my own definition:

> Culture is an information pool that emerges when members of a community attempt to make sense of the world and each other as they struggle and collaborate with each other to get what they want and need (e.g., food, sex, power, acceptance, etc.). Because individuals construct their conceptions of the world from their own experiences and for their own motivations, their understandings vary from one another depending on the characteristics of the individuals, the nature of the domain learned, and the social situations in which learning takes place. [modified from (*Boster 1986*)]

This definition shares Goodenough's insistence that culture is learned knowledge, but adds that culture is (usually) useful and that individuals vary in their knowledge, and emphasizes that an ethnography should describe and explain the social distribution of knowledge.

[5] The field was generally called "ethnoscience" at the time and its practitioners "ethnoscientists." The term "cognitive anthropology" was not coined until the publication of Tyler's (1969) anthology. Ron Rohner, a fellow graduate student at Stanford in the early 1960s, takes credit for suggesting the title to Tyler.

their task to an exploration of the cognitive organization of lexicons. Frake (1962, p.74) offered a succinct justification for this narrowing of scope:

> To the extent that cognitive coding tends to be linguistic and tends to be efficient, the study of the referential use of standard, readily elicitable linguistic responses – or *terms* – should provide a fruitful beginning point for mapping a cognitive system. And with verbal behavior we know how to begin.

The principal tool used to study the organization of terms in a language was componential analysis. This can be defined as a procedure for discovering the dimensions of meaning underlying a domain and mapping values on these dimensions (or components) onto terms in a lexicon. The procedure was directly analogous to those which had proved so useful to linguists in discovering phonemes and morphemes; one finds pairs or sets of terms that contrast with each other (i.e., are included in the same superordinate category), one attributes everything that is shared in meaning to membership in the superordinate category, and finally one discovers the attributes that distinguish among the members of the contrast set[6]. While Frake's (1962) description of the method allowed it to be applied to taxonomies (nested contrast sets related by inclusion), Goodenough (1956) and Lounsbury (1956) further restricted the application of the method to those lexicons that formed paradigms[7], in which every (or nearly every) combination of feature values corresponded to a distinct term.

The most perfect example of the application of componential analysis was Conklin's (1962) analysis of Hanunóo pronouns shown in Figure 1. The eight pronouns corresponded to every possible combination of three binary feature contrasts – inclusion and exclusion of speaker ($S,\bar{S}$), inclusion and exclusion of hearer ($H,\bar{H}$), and minimal and non-minimal membership ($M,\bar{M}$) – and the whole set of pronouns could be represented as the vertices of a cube, whose depth, width, and height corresponded to the three feature dimensions.

Unfortunately, examples of such heart-squeezing beauty and elegance were rare. Beyond a few applications of componential analysis to pronominal systems and kinship systems, there were few domains organized as paradigms. And even kin terminologies fit awkwardly into strict paradigms. One reason for the rarity of paradigms can be seen by comparing the number of terms required to label every possible combination of feature values with the number required to label the feature values themselves, as shown in Table 1.

[6] The property of language exploited in these methods was its capacity to make near infinite use of finite means. At every level of the analysis of language (phonology, morphology, semantics, and syntax), one finds a finite set of elements (phonemes, morphemes, grammatical classes) combined through a finite set of rules (phonological rules, morphological rules, syntactic rules) to generate a much greater number of possible forms. Linguists (and cognitive anthropologists) use this property of language to work backwards, systematically comparing minimally different complex units (sounds, words, meanings) to discover the underlying universe of elements and the rules for combining them.

[7] This use of "paradigm" is in the same sense as "verb paradigm," the pattern of verb inflections which often take different forms according to whether speaker and/or hearer, etc., are included in reference, etc. It has nothing to do with Kuhn's (1962) use of the word "paradigm" in the sense of a scientific framework.

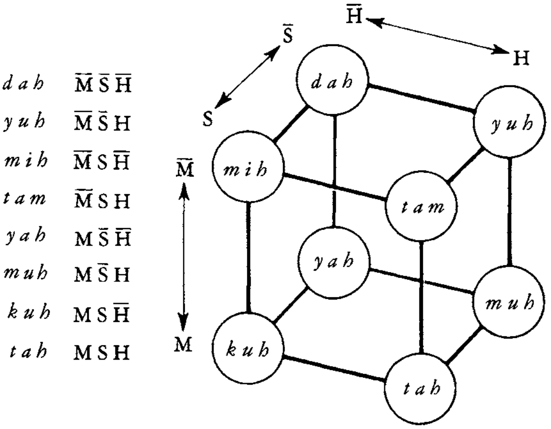

Fig. 1. Componential analysis of Hanunóo pronouns. From Conklin (1962). Reprinted with permission.

Table 1
Lexical cost of a perfect paradigm

Number of binary feature dimensions (*N*)	Number of terms to label to combinations of feature values (2^N)	Number of terms needed to label feature values ($2N$)	Lexical cost of labeling combinations of feature values relative to labeling feature values directly ($2^N/2N$)
1	2	2	1.00
2	4	4	1.00
3	8	6	1.33
4	16	8	2.00
5	32	10	3.20
6	64	12	5.33

If N is the number of binary feature dimensions, the number of terms required to label the unique combinations of features is 2^N, while the number of terms required to label the distinct feature values is $2N$. Thus, as the number of feature dimensions increases, the lexical cost of labeling the terminal nodes of a perfect paradigm relative to labeling the feature values (the ratio $2^N/2N$) increases exponentially. A feature key of a perfect paradigm [using the conventions of Kay (1966)] is shown in Figure 2.

Feature dimensions are indicated by letters (e.g., "a," "b," "c," ...) and feature values are indicated by subscripts (e.g., "a_1," "a_2," "b_1," "b_2," "c_1," "c_2," ...). (For example, "a" could be used to signify the feature dimension of *sex*, and "a_1," "a_2" used to signify the feature values of *male* and *female*, respectively.) With two dimensions, one needs just as many terms (four) to label the unique combinations of features as to label

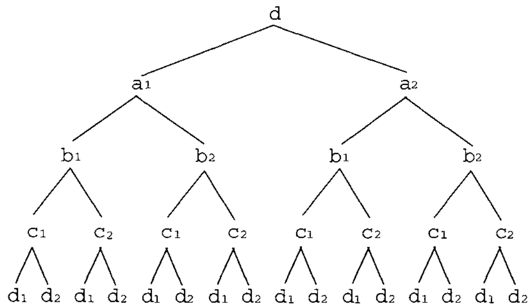

Fig. 2. Feature key of a perfect paradigm. Adapted from Figure 2a, of Kay (1996).

the distinct feature values. With three dimensions, one needs a third again as many terms to label the unique combinations as to label the feature values, with four dimensions, twice as many, and by the time one gets to six feature dimensions, it requires more than five times as many terms to label the unique combinations of feature values as to label the unique features alone. Six feature dimensions is the theoretical limit according to Wallace (1964), corresponding to Miller's (1956) "seven plus or minus two" constraint on short-term memory. If it were not for this constraint, the disparity would be 2,048 times as many terms with 16 feature dimensions, and so on. No wonder paradigms are rare; very frequent reference to the unique combinations of feature values (as in pronominal systems) is required to justify the lexical overhead.

However, a more profound reason for the rarity of paradigmatic organization of lexicons may be that the very independence of feature dimensions that would allow them to be represented in neat nested boxes deprives them of the "entitivity" necessary to be recognized as classes of things that require their own labels. "Entitivity" is a word coined and defined by Campbell (1958, p.17) as "the degree of being entitative. The degree of having the nature of an entity, of having real existence." Campbell was addressing the question of when one calls aggregates of human beings a social group. In order to do so, the group had to have some degree of entitivity, so as to distinguish accidental agglomerations of people (e.g., the passengers waiting at a bus stop) from members of more "real" groups (e.g., the lifetime members of a professional organization, the members of a nuclear family). Campbell listed five criteria: common fate, similarity, proximity, resistance to intrusion, and internal diffusion. Campbell was building on insights by Herbert Spencer (1897), who regarded a living body as an entity *par excellence.* Most of Campbell's criteria were borrowed from those that Gestalt psychologists used to decide whether something constituted a gestalt. I would add two more criteria to his list when deciding whether a collection of things constitutes a set of entities worth classifying: "do the features that characterize the items vary independently or interdependently?" and "do the features vary continuously or are the feature values discrete?" Both interdependence of feature correlation and discreteness of feature values contribute to the "entitivity" of items in a domain.

Consider soils. They vary continuously in texture from very fine clays through silts and loams to coarse sands and gravels. The variation in texture is largely independent of the equally continuous variation in color, as indicated by the huge number and subtle gradations of colors found in the Munsell soil color book. If every single possible combination of color and texture of soil were categorized as a distinct kind of soil and given its own label, our folk soil classification would dwarf our folk biological classification. This does not happen. Folk systems tend to have a few terms like "clay," "silt," "loam," and "sand" that describe variation in texture and a few other terms like "muddy," "parched," and "rocky" to describe degrees of saturation with water or stones. Folk systems have the possibility of describing the colors of soils just as they can describe the color of almost anything else. However, they do not have nearly the number of terms necessary to distinctly name all possible combinations of variation in texture and color. The K'ekchi Maya corn farmers I worked with some years ago classified soils principally into just two classes: *chabil ch'och* ('good earth') and *ma us aj' ch'och* ('bad earth'), distinguished by their suitability for corn farming. Even soil scientists rarely use the named classification of soils, instead describing the horizons of the soil profile by their color (determined by consulting the Munsell soil book) and texture when surveying soils. Although there is sufficient variation in the characteristics of soils to allow many different types of soil to be differentiated, because of the continuity and independence of the feature dimensions, they do not make for "discontinuities in nature calling out to be named." They also lack other aspects of entitivity described by Campbell, especially the lack of any clear boundaries between one soil type and another and the absence of any appreciable resistance to intrusion.

The greatest degree of entitivity, using Campbell's amended criteria, is found among living things, as Spencer intuited. Biological evolution is a wonderful structure-generating engine. Each speciation event distinguishes a pair of species on at least one and usually several features. Once they are no longer part of the same breeding population, the two new species may respond to different sorts of selection pressures, so that the granddaughter species cleaved by subsequent speciation events are usually not distinguished on the same features as the other granddaughter species. The feature key (again, using the conventions of Kay, 1966) representing this process is shown in Figure 3, which not coincidentally, is also the feature key of a perfect tree.

In this case, there is a lexical economy in naming the unique combinations of feature values rather than the feature values themselves, because the features multiply faster than their unique combinations do, as shown in Table 2. Again, if N is the number of feature dimensions, the number of terms required to label the unique combinations of features is $N+1$, while the number of terms required to label the distinct feature values themselves is $2N$. Thus, as the number of feature dimensions increases, the lexical cost of labeling the terminal nodes of a perfect tree relative to labeling the feature values (the ratio $(N+1)/2N$) decreases asymptotically to ½.

Another difference is that in a paradigm there is no intrinsic ordering of the feature dimensions, while in a tree there is an intrinsic order. Kay (1966) referred to this difference as the contrast between the simultaneous application of feature dimensions in a

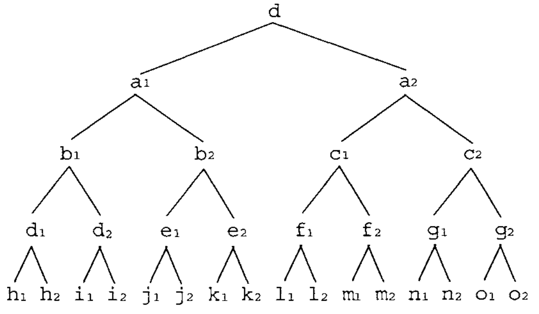

Fig. 3. Feature key of a perfect tree. Adapted from Figure 2b, of Kay (1996).

Table 2
Lexical cost of a perfect tree

Number of binary feature dimensions (N)	Number of terms to label to combinations of feature values ($N + 1$)	Number of terms needed to label feature values ($2N$)	Lexical cost of labeling combinations of feature values relative to labeling feature values directly ($(N + 1)/2N$)
1	2	2	1.00
3	4	6	0.67
7	8	14	0.57
15	16	30	0.53
31	32	62	0.52
63	64	126	0.51

paradigm versus the sequential application of feature dimensions in a tree. Because a particular feature dimension applies at one and only one node of a tree, labeling the superordinate nodes is meaningful in a tree in a way that it is not in a paradigm. If the superordinate nodes of a tree are labeled, it is called a taxonomy. A third difference between paradigms and trees is their difference in feature redundancy. The excess of distinguishing features over the number of categories to be distinguished in trees makes it much more likely that the combinations will be "chunked" into an integrated gestalt, something that the feature independence of paradigms does not afford. It is the interdependence of the features of trees that Rosch and Mervis (1975) are referring to with their phrase "correlational structure." The greater lexical economy in labeling unique combinations of feature values, the greater entitivity of the items in the domains, and the more memorable "gestalts" of feature combinations make trees and taxonomies much more common modes of semantically organizing lexicons than paradigms. This meant that componential analysis was soon displaced by other methods of exploring the semantic organization of lexicons as the principal tools of cognitive anthropologists, especially by those based on taxonomies [e.g., Metzger and Williams (1966)] or open-ended semantic networks [e.g., Frake (1964)] and by methods borrowed from psychology [e.g., Romney and D'Andrade (1964)].

4. Kinship terminologies

An early demonstration of the way in which the choice of data analysis method could affect the representations of category structure came in the analysis of kinship terminologies. As a way of showing off the power of their methods, various early cognitive anthropologists analyzed the structure of American English kin terms [e.g., Wallace and Atkins (1960), Romney and D'Andrade (1964),Goodenough (1965)]. As Schneider (1965) has pointed out, using American kinship as the arena for the comparison of methods allowed readers who were native speakers of American English to criticize the analyses with their own native intuitions, unlike their helplessness in the face of descriptions of the exotic. Two theoretical questions came to the fore in this exchange: "how do you choose between alternative analyses?" and "how do you demonstrate the psychological validity of a particular analysis?" Both questions stemmed from the sense that an analysis of a semantic domain should reflect the cognition of the informants, not the investigator.

Wallace and Atkins (1960) described the method of componential analysis of kinship terminologies propounded by Goodenough (1956) and Lounsbury (1956) as consisting of five steps: (1) recording a complete set (or defined subset) of kin terms; (2) defining the terms by their kin types (e.g., MoFa, MoBrSo); (3) identifying two or more feature dimensions whose values or components are signified by the terms; (4) defining each term by its feature values; and (5) stating the semantic structure of the whole set of kin terms. Starting with a defined subset of English kin terms (the consanguineal core terms), they end up with the structure shown in Table 3.

In their discussion, Wallace and Atkins recognize that alternate analyses are possible and that one should prefer the analysis that is "the most proximate to psychological reality" (p. 78). However, the authors recognize the difficulty of developing methods to establish psychological validity and do not actually present any themselves.

Table 3
A componential analysis of American English kin terms

	lineal		colineal		ablineal	
	male	female	Male	female	male	female
+2	*grandfather*	*grandmother*	*uncle*	*aunt*		
+1	*father*	*father*				
0	[ego]		*brother*	*sister*		*cousin*
−1	*son*	*daughter*	*nephew*	*niece*		
−2	*grandson*	*granddaughter*				

Adapted from Figure 1 of Wallace and Atkins (1962).

Goodenough (1965) later analyzed a greater range of kin terms (including step- and in-law terms) and solved the problem of psychological validity to his own satisfaction by using himself as his own informant and rejecting analyses that "just didn't sit right." His methods and resulting analysis, however, are substantially similar to those of Wallace and Atkins (1960).

In contrast, Romney and D'Andrade (1964) present a radically different approach both to the problem of producing an analysis and to that of establishing its psychological validity. They begin with the same set of consanguineal core kin terms used by Wallace and Atkins (1960), but define the terms using a special notation to indicate the kin paths between ego and various alters, rather than the usual kin type notation. In this notation, "m" indicates a male; "f" a female; "a" someone of either sex; "0" a sibling link; "+" a parent link; "−" a child link; "=" a marriage link; and "/" is used to enclose expressions that can be read as reciprocals. Thus, a mother's brother's son would be represented as "a+f0m−m" rather than "MoBrSo." The advantage of this notation is that it allows the authors to describe a set of rules that reduce to a single expression all the paths linking ego to various alters indicated by a kin term. Romney and D'Andrade's (1964) procedure produces the structure shown in Table 4.

Note that the principal differences from Wallace and Atkins (1960) are that there are only two degrees of lineality (direct and collateral) rather than three (lineal, colineal, and ablineal) and that generational terms that are reciprocals of one another (e.g., *grandmother*, *granddaughter*; *mother*, *daughter*) are treated as sharing the feature value of being one or two generations away from ego, rather than distinguishing kin terms by their absolute generation[8].

Romney and D'Andrade's (1964) second major innovation was to describe a set of procedures to compare the psychological validity of their own model with that of Wallace and Atkins (1960). The critical data for the comparison were elicited using a triad test in which they presented all possible triplets of a set of eight male kin terms to 116 high school students and asked of each triad, "Which term is most different from

Table 4
An algebraic analysis of American English kin terms

		direct		collateral	
		male	female	male	female
two generations	+	*grandfather*	*grandmother*		
different	−	*grandson*	*granddaughter*		
one generation	+	*father*	*mother*	*uncle*	*aunt*
different	−	*son*	*daughter*	*nephew*	*niece*
same generation		*brother*	*Sister*	*cousin*	

Adapted from Figure 3 of Romney and D'Andrade (1964).

[8] Same-generation terms are reciprocals of themselves (e.g., I am my brothers' brother and my cousins' cousin.)

the other two?" They found that terms that were reciprocals of one another (e.g., *grandson* and *grandfather*) were judged as more similar than were terms that were closer in absolute number of generations (e.g., *grandfather* and *father*), as predicted by their own model and not by Wallace and Atkins (1960). Furthermore, the three dimensions of a multidimensional scaling of the similarity judgments corresponded to the three dimensions of Romney and D'Andrade's (1964) model: reciprocity, generational difference, and lineality. They concluded that their model has greater psychological validity than the alternative model described by Wallace and Atkins (1960).

Little turns on this particular conclusion, for it hangs on a slender thread. True, when given the triad *grandfather, father*, and *grandson*, most American English speakers will judge *father* to be the most different from the other two. However, a subtle change in wording produces a contrary result; if given the triad *my grandfather, my father*, and *my grandson*, most will pick *my grandson* as the most different, as predicted by Wallace and Atkins (1960). Notice that the change in wording has this effect because it shifts attention away from the relationship 'grandfather' or 'father' to a particular individual who is someone's grandfather or father. The mature elders who are fathers and grandfathers are much more similar to each other than the callow youth who is their son or grandson, even if the relationship 'grandfather' is more similar to its reciprocal 'grandson' than it is to the relationship 'father' or 'son.' This shift of attention from relationships to individuals better matches the typological assumptions of componential analysis[9].

More significant is Romney and D'Andrade's development of methods for establishing the psychological validity of alternative models of semantic structure. But here again, although this was an important step forward at the time, in retrospect, it has had little lasting effect on the field, because it was not followed by a flood of papers that developed a model using one set of methods and tested its validity with another set. Instead, the psychological and statistical methods that Romney and D'Andrade (1964) introduced to test for psychological validity soon displaced the primarily linguistic methods that had hitherto been used to develop the models. Investigators began eliciting similarity judgments to directly build their models of semantic structure and dispensed with componential analysis and related methods altogether. Thus, multidimensional scalings and cluster diagrams of similarity judgments replaced the keys and boxes of taxonomies and paradigms as figures in cognitive anthropological articles. In the process, the notion of testing models of semantic structure for their psychological validity with an independently collected set of data was lost.

There is also the sad fact that interest in studying kinship and kin terminologies has largely withered among anthropologists and linguists. Hardly anyone noticed a few years later when Ellen Woolford (1984) offered the definitive analysis of kinship terminological systems. Woolford combined elements of three different approaches (componential

[9] By "typological assumptions," I mean the assumptions that the terms in a lexicon label a set of items in a semantic domain and that each of the items can be distinguished from the others by some small set of distinguishing features. In the case of kinship terminologies, the terms refer to relations between items (the individual humans); they are not distinguished by features of the items in isolation.

analysis, extension rule analysis, and relational analysis) in a clear and elegant generative model that predicts a number of implicational universals (e.g., "if you refer to the same sex siblings of your parents with the same terms you use for your parents, you will refer to their children with the same terms you use for your siblings"). The field has not died altogether. Dwight Read and his collaborators continue to analyze kin terminologies with (kinship analysis expert system) (KAES) [(Read and Behrens, 1990)]. However, one needs a test similar to that performed by Romney and D'Andrade (1964) to decide whether KAES produces more psychologically real models than its predecessors.

I believe that the most important contribution of Romney and D'Andrade's (1964) work has been little appreciated. The reason why their procedure produced an analysis that was superior to that produced by componential analysis is that the algebraic analysis of relationship paths is truer to the logic of the domain of kinship than the typological assumptions underlying componential analysis. That is, although the term *niece* picks out of the world a small set of particular (and charming) human beings for me, the term really does not refer to particular humans but to a relationship between humans. I call them *nieces*, they call me *uncle*, but all of us bear a variety of kin relationships to others. In other words, I am not an 'uncle' in general, but only an 'uncle' to my nieces and nephews. The algebraic notation and reduction rules developed by Romney and D'Andrade (1964) capture the logic of these relationships better than any typological method. It is a shame that the authors demonstrated the approach by analyzing American English kin terms, for its real power is shown off best in making the exotic transparent, as D'Andrade (1995, pp.20–28) shows in analyzing Bellah's (1952, p.71) data on Chiricahua kin terms[10]. Using Romney and D'Andrade's (1964) notation and reduction rules, one can derive a single expression `/a+m(+a)0(a-)a/` which represents the 40 relationship paths from ego to various alters that are referred to by the Chiricahua kin term *cìdèèdèè*. (Similar reductions can be accomplished for the other Chiricahua kin terms.) Romney and D'Andrade (1964) heralded what became a common pattern in the development of cognitive anthropology. Our understanding of particular domains was often improved when the methods and assumptions of the analysis better matched the nature of the domain being studied[11].

Another work of this period, Nerlove and Romney (1967), also had far-reaching implications, not so much for the importance of the domain studied (sibling terminologies) as for its demonstration that the methods that had been used to explore the internal logic of a lexicon in a single language could be used to systematically compare semantic systems. The first step was to describe a universe of logical possibilities. There are eight possible kin types given by all possible combinations of sex of ego (x = male, o = female), sex of alter (B = male, Z = female), and age difference of alter and ego (e = elder, y = younger) and there are one to eight possible kin terms to refer to these

[10] Shortly after the publication of Romney and D'Andrade (1964), Romney (1965) applied the method to the analysis of Kalmuk Mongol kinship terminology but this article never became as well known as the earlier work.

[11] John Atkins' (1974) GRAFIK metalanguage was arguably a superior algebraic notation system for kinship.

kin types, yielding a potential universe of 4,140 possible ways of assigning kin types to categories labeled by the kin terms. (For example, in English, the kin types `xZe`, `oZe`, `xZy`, and `oZy` are all called *sister* and the kin types `xBe`, `oBe`, `xBy`, and `oBy` are all called *brother*. In contrast, in Awajún (Aguaruna Jívaro), the kin types `oZe` and `oZy` are called *kai*, the kin types `xBe` and `xBy` are called *yatsu*, and the kin types `xZe`, `xZy`, `oBe`, and `oBy` are called *uma*.) Nerlove and Romney's next step was to characterize the sibling terminologies found in a sample of 245 societies using this notation system.

Out of the 4,140 conceivable ways of categorizing the kin types, Nerlove and Romney (1967) found only 12 regularly occurring types with a handful of outliers in their sample. Thus, the cross-cultural variation in sibling terminologies is both much greater than would be guessed from English alone (only 10% of societies distinguish siblings according to absolute sex alone) and much smaller than the range of possible variation. Nerlove and Romney explained the constraints on the variation in terms of universal cognitive and social factors (e.g., disjunctive categories are rare compared with conjunctive or relational categories; categories in which sex of ego is a primary distinction are nonexistent), and they explained the variation in terms of social variables (e.g., sibling categories in which relative sex of ego and alter is a primary distinction are most common in societies that have brother–sister avoidance and long postpartum sex taboos). This article was an important development because it used explicit formal methods to explain patterns of cross-cultural similarities and differences. Although this work has been singled out by Shweder and Bourne (1984, pp.160–161) as exemplifying a mode of finding cultural universals by excessively circumscribing the scope of study, such circumscription to sibling terms, and not kinship terms in general, was probably necessary to be able to identify a coherent and tractable comparative project. Romney and D'Andrade (1964) also represented a rejection of the relativism of the early cognitive anthropologists and a return to Murdoch's systematic cross-cultural comparison. This became the model for later cross-cultural comparison of categorization in color, ethnobiology, and other domains as discussed below.

Before we turn our attention to the development of an understanding of color classification by cognitive anthropologists, it is useful to review the steps used by Nerlove and Romney (1967) that were emulated in later investigations of patterns of cross-cultural similarities and differences. First, one picks a domain and examines the structure of the feature space. Second, one isolates comparable units of analysis. Third, one compares the systems and describes patterns of similarities and differences. Finally, one attempts to explain the similarities in terms of shared constraints, and the differences in terms of factors that differentially affect the societies. Unfortunately, cognitive anthropologists did not always get around to performing this last step.

5. Color classification

Color is another domain in which the choice of method (especially the choice of how the units of analysis were defined) affected the representation of category structure.

Shifts in theoretical interpretation were also driven by the collection and analysis of an ever-increasing quantity and quality of data. Early studies treated color classification as an example of extreme cultural relativism. In Roger Brown's textbook, *Social Psychology*, the color spectrum was represented as a continuum that is segmented by different languages in arbitrarily different ways (1965, pp. 315–316). The empirical reports that were available supported the extreme relativist position. For example, Conklin's (1955) study of Hanunóo color categories described them as encoding features including surface texture and sheen, not just hue, brightness, and saturation. Furthermore, one would expect cultural relativism in color categorization given the nature of the domain. Colors vary continuously from one another in hue, brightness, and saturation. The continuity and independence of the feature dimensions deprive the domain of entitivity. If there is no intrinsic structure to the color space to guide their category boundaries, nothing would prevent speakers of a different language from cutting the continuum up in a way that is arbitrarily different from the way English does it.

But this view was not at all correct. As Paul Kay in Tahiti and Brent Berlin in Tenejapa observed from their experiences doing fieldwork, it was not nearly as difficult to learn the meanings of the color terms in Tahitian or Tzeltal as would be implied by an extreme relativist position. They addressed the question in a perfectly sensible way for assistant professors at Berkeley: they organized a graduate seminar and sent their students throughout the San Francisco Bay area with Munsell color charts to see how the color terms in different languages mapped to a common set of referents. What they found won them fame and tenure, and established what they later called the Universals and Evolution approach to color classification. The breakthrough was achieved by defining color categories by their best examples, because, although there was considerable interinformant variation on the placement of the boundaries between color categories (even with the use of a stimulus array that encouraged well-bounded categories), there was considerably more agreement both within and between languages on the choices of the best examples, or foci, of the categories. Berlin and Kay (1969) found that there were only 11 basic color categories in the 20 languages they were able to investigate, with foci in black, white, red, yellow, green, blue, brown, gray, pink, orange, and purple. This confirmed their initial hunch. More surprisingly, the color categories came in a limited number of combinations. If there were two categories, their foci were in black and white; if there were three categories, they focused in black, white, and red; if four categories, black, white, red, and yellow, or black, white, red, and green; if five categories, black, white, red, yellow, and green; if six categories, blue was added; if seven, brown was added; and if eight or more categories, gray, pink, orange, and purple were added in no particular order. This pattern was interpreted as the sequential encoding of foci, as though the speakers of the various languages had a push-down stack of color foci (like the floating stacks of plates found in some cafeteria lines), with white and black at the top (at stage I), red next (stage II), yellow or green next (stage IIIa or IIIb), green or yellow next (stage IVa or IVb), followed by blue (stage V), brown (stage VI), and a random shuffle of gray, pink, orange, and purple foci at the bottom (stage VII). Berlin and Kay (1969) had to guess what a two-category system would look like because none of the languages they had access to had only two terms, but their framework accommodated all of the systems that they and their students were either able to investigate directly or surmise from dictionary entries.

But this model was not quite right. When Eleanor Rosch (then Heider) later investigated a two-term color classification system among the Dani of Irian Jaya, she found that the 'white' category actually focused in white, yellow, or red and was better described as a warm/light category, while the 'black' category actually focused in black, green, or blue and was better described as a cool/dark category [Heider (1972a,b)]. Similarly, Berlin found that the 'green' term in Tzeltal Mayan and other stage IV systems could focus in either green or blue and was better described as a [GRUE] category. To accommodate these new data, Kay and McDaniel (1978) reinterpreted the evolutionary[12] sequence (at least until stage V) as the successive differentiation of existing color categories. Thus, a two-term system at stage I was interpreted as having two categories of unique hue points: white, red, and yellow versus green, blue, and black. With three categories, white was split from red and yellow; with four categories, either yellow was split from red or green and blue from black; with five categories, red and yellow were split apart, and green and blue were split from black; and with six categories, all six of the unique hue points were split into separate categories. Secondary basic color terms were interpreted as the intersection of unique hue points: brown as the intersection of yellow and black; gray as the intersection of white and black; pink as the intersection of white and red; orange as the intersection of red and yellow; and purple as the intersection of red and blue. Languages could have categories defined by the intersection of unique hue points as soon as the constituent hue points were in their own separate categories. This helped account for the "premature gray" found in some Amerindian languages.

Kay and McDaniel (1978) also argued that the unique hue points and the cross-cultural universals they gave rise to had their origin in the neurophysiology of color vision. They based their interpretation on DeValois, Abramov and Jacobs (1966) research on the lateral geniculate nucleii (LGN) in the hypothalamus of Rhesus macaques which had revealed three families of neurons. Two families were opponent processes: a red–green channel (excited by red light and inhibited by green light, or vice versa) and a yellow–blue channel (excited by yellow light and inhibited by blue light or vice versa). The third family was a white–black channel which responds to brightness levels independently of the other two channels. This physiological account helped explain why there might be universals in the classification of what seemed a structureless domain. The structure is imposed by the neurophysiology of color vision.

But this schema was still not quite right. One problem was that anomalous composite categories were found, which joined yellow, green, and blue; yellow and green; white and yellow; or blue and black. None of these combinations could be derived from the successive differentiation of color categories as described by Kay and McDaniel (1978). Furthermore, subsequent research had revealed that the neurophysiological opponent process system would actually put the unique hue points in the wrong places. The true

[12] The term "evolutionary" in this context refers to lexical, not biological, evolution.

axes of the system are closer to cherry–teal and chartreuse–violet than they are to red–green and yellow–blue [Jameson and D'Andrade (1997)]. Sadly, this left the universals in color classification without a clear physiological explanation. Repairs to the framework were made by Kay, Berlin and Merrifield (1991), which provided the first report of the World Color Survey (WCS). The WCS investigated color naming in 111 languages, interviewing roughly 25 native speakers of each language about their color names for each of 330 color chips, and their choices of the best examples of each basic color term of their language. The WCS represented an enormous improvement in both the methods and the quantity of data over what was available to Berlin and Kay (1969). The 330 color chips were presented to informants one by one in random order, thereby overcoming a bias toward the recognition of contiguous color categories that was inherent in the use of the original Munsell grid of color chips arranged in 8 rows and 40 columns (plus the gray scale). Many more speakers of each language were interviewed, thereby allowing a genuine comparison of the magnitudes of intra- and interlanguage variation. And finally, nearly six times as many languages were surveyed, most of which were spoken by indigenous groups in Africa, Central and South America, Papua New Guinea, and smaller numbers of indigenous groups from North America, India, Indonesia, the Philippines, and Australia, all areas where the linguists of the Summer Institute of Linguistics were working. Most of the languages sampled had early stage color systems. Because so many more languages were studied beyond the initial set of 20 in Berlin and Kay (1969), it was practically inevitable that the researchers would encounter greater variation than they had initially. To account for this greater variation, they formulated the "composite category rule" which posited a network of unique hue points in which white, red, and green were each independently linked to yellow, blue was linked to green, and black linked to blue as shown in Figure 4. Their composite category rule allowed for the existence of any composite category that joined linked hue points which did not cross yellow. This rule allowed all of the composite categories permitted by Kay and McDaniel (1978) (red, white, yellow; green, blue, black; red, yellow; green, blue) plus a number of additional composite categories which had previously been treated as anomalous (yellow, green, blue; white, yellow; yellow, green; and blue, black). Their rule also allowed a composite category that had not yet been found (yellow, green, blue, black).

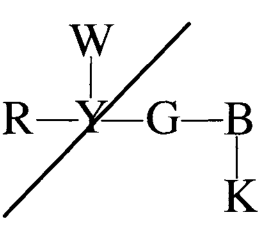

Fig 4. Network of unique hue points in the composite category rule. Adapted from Figure 2 in Kay, et al. (1991).

But even this formulation was not quite right. The problem was that the solution to the existence of anomalous categories was awkward; the composite category rule was an arbitrary statement of permitted categories and offered no statement of its own logic. It was a step backwards, and one that Kay et al. (1997) and Kay and Maffi (1999) rectified. The rectification was a partial reversion to the architecture of the original Berlin and Kay (1969) evolutionary sequence, but instead of recognizing just two evolutionary trajectories, they now recognized five. The five trajectories were interpreted as generated by four principles: partition (lexicons tend to partition items in culturally significant domains into exhaustive and mutually exclusive categories); black and white (color lexicons distinguish black and white); warm and cool (color lexicons distinguish the warm primaries from the cool primaries); and red (color lexicons distinguish red from other colors). Mysteriously, the yellow-white composite category found in Waorani, an Amazonian Amerindian language, was no longer a permitted composite category, but this was a small price to pay for the much greater clarity and coherence of the framework.

But this picture is unlikely to be quite right, given the long history of revisions of the framework. Cook,Kay and Regier (this volume) offer the latest results of the analysis of the WCS. Still, I believe this history is a heartening one. Although the most recent statement of the evolutionary sequence is not nearly as simple as the original one, it is based on much better evidence from about six times as many languages as the original. The revised framework manages to preserve the most important element of the Universals and Evolution tradition: the notion that underlying the considerable cultural variation in how languages categorize colors, there is nevertheless evidence of a universal pattern.

6. Ethnobiology

Ethnobiology, the study of the folk classification of plants and animals, is a third domain that illustrates the impact of changes in methods of data collection and analysis on our understanding of the nature and structure of semantic domains. In particular, two works by Berlin and colleagues illustrate how the shift from a relativist to a universalist interpretation came about mainly by virtue of a shift in how the units of analysis were defined. Berlin, Breedlove and Raven (1966, p. 273) clearly and simply defined a folk specific category as "any taxon which includes no other taxa." This definition had the advantage of unambiguously specifying the units of analysis as the set of all terminal taxa – categories not further subdivided. They then examined the ways in which the folk specific taxa corresponded to scientific species. Forty-one percent of their sample of 200 Tzeltal folk specifics were underdifferentiated with respect to scientific species (a single Tzeltal folk specific included more than one scientific species), 34 percent were in a one-to-one correspondence with scientific species, and 25 percent were overdifferentiated with respect to scientific species. They concluded that "Tzeltal specifics clearly do not correspond in any predictable way with botanical species" (1966, p. 273).

Berlin et al. explained the variation in the way folk specifics map to scientific species by appealing to the cultural significance of the plants. Culturally significant plants such as food crops and other useful plants were likely to be overdifferentiated, economically unimportant species were likely to be underdifferentiated, and introduced species were more likely to be in one-to-one correspondence. Furthermore, those folk specifics that are overdifferentiated with respect to scientific species often had the form of "attributive + head" where the head term referred to a category in one-to-one correspondence with a scientific species (e.g., terms like *sweet corn*, *pop corn*, *dent corn*; *corn* = *Zea mays*).

A short time later, Berlin et al. had a change of heart when they looked at their data in a different way. This is an unusual case [as noted by Gould (1980)] in which the same team of researchers investigating the same people's knowledge of the plant world completely reversed themselves, and shifted from a relativist to a universalist interpretation of essentially the same data. The critical shift came in how they defined a folk species. Berlin, Breedlove and Raven (1973, p. 214) state this change of heart:

> The richness and diversity of man's variant classifications of experience have often led ethnographers to emphasize the differences between cultural systems of knowledge.... There are a number of strikingly regular structural principles of biological classification which are quite general.

Their principles were that ethnobiological domains in all languages are organized taxonomically; that taxa are grouped into a small number of ranks (unique beginner, life-form, generic, specific, varietal) that are hierarchically arranged; taxa of the same rank usually occur at same level of the taxonomy and are mutually exclusive. The most inclusive rank, the unique beginner, is often unlabeled; life-forms are immediately included in the unique beginner, are few in number (5–10), include most lower taxa, and are polytypic (e.g., *tree*, *vine*, *bird*). Categories at the folk generic rank are the most numerous, although most are included in life-forms, some are unaffiliated, but all are highly salient and commonly referred to, learned early by children, and labeled by primary lexemes (e.g., *oak*, *pine*, *robin*, *cactus*, *banana*). Specific and varietal categories are fewer in number, occur in small contrast sets, are often culturally important, distinguished by few features, and labeled by secondary lexemes (e.g., *live oak*, *valley oak*, *Ponderosa pine*, *Monterey pine*, *butter lima bean*, *baby lima bean*). Finally, intermediate categories are included in life-form categories and include generics, are often unlabeled, and are very rare (e.g., *evergreen*, *deciduous*, *songbird*). The advantage of these principles of folk biological classification is that they allowed one to recognize a much greater degree of correspondence between folk and scientific systems of classification, and also much higher degrees of structural similarity among the folk systems themselves. For example, the "attributive plus head" term found among many overdifferentiated folk specifics in the previous (1966) scheme, remained folk specifics in the new system, but the head term without any attributive was recognized as a folk generic, and thus more comparable to the categories that were in one-to-one correspondence or underdifferentiated in the earlier scheme. The cost of this insight was much greater overhead in complicated rules to distinguish between categories at the different ranks and between the different types of labels of those categories.

Berlin et al. (1973) use the form of the terms for taxa as an important clue to their ranks. They distinguish primary lexemes which label life-form and generic rank taxa from secondary lexemes, which label specific and varietal rank taxa. Primary lexemes are mostly single-word expressions that are "semantically unitary and linguistically distinct" (p. 217). They are either unanalyzable (e.g., *pine*, *oak*, *tree*) or analyzable. If analyzable, they are either productive (in which case, one of the constituents refers to the including class, but not all members of the contrast set share the superordinate label, e.g., *plane tree*, *pipe vine*, *lead tree*) or unproductive (in which case, no constituent marks the superordinate class, e.g., *beggar-tick*, *cat-tail*, *poison oak*, *jack-in-the-pulpit*). In contrast, secondary lexemes usually have the form "attributive + head," in which case one of the constituents refers to the including class and all the members of contrast set share same superordinate constituent (e.g., *white pine*, *jack oak*, *swamp beggar-tick*). The tricky distinction is between productive analyzable primary lexemes and secondary lexemes. Both have the form "attributive + head," where the head term refers to the including class. One must review the other members of the contrast set to see if the inclusion of the head term is obligatory for all the contrasting terms.

The payoff of these fine distinctions is that one could tell the rank of a category from its label, something that was critical in order for Berlin et al.'s (1973) theory of ranks to work. Their nomenclatural principles worked as follows. If a taxon is labeled by a primary lexeme and either is terminal or includes taxa marked by secondary lexemes, it is a generic. If a taxon is labeled by a primary lexeme and includes taxa marked by primary lexemes, it is a life-form[13]. If a taxon is labeled by a secondary lexeme, is immediately included in a taxon labeled by a primary lexeme, and either is terminal or includes taxa marked by secondary lexemes, it is a folk specific. Finally, if a taxon is labeled by a secondary lexeme, is terminal, and is immediately included in a taxon labeled by a secondary lexeme, it is a varietal.

This was an important advance over their earlier work, for it allowed Berlin et al. (1973) to recognize structural similarities among a large number of folk systems of biological classification. Unfortunately, there were also a number of problems with the approach. First, there was a problem with the psychological status of the taxonomic representation of folk biological knowledge in the first place. If your informant told you that a *white pine* was a kind of a *pine*, that a *pine* was a kind of an *evergreen* and that *evergreens* were *trees*, did that mean that they really had the whole structure in their head? Randall's (1976) answer was "No!" He argued that one could elicit a very similar chain from an informant concerning what animals eat which other animals (e.g., chickens eat worms, humans eat chickens, worms eat humans), but one could not conclude after the interview that the informant had a concept of trophic levels. Indeed, sometimes the inclusion relations are not transitive – a *scrub* oak is a kind of *oak,* and an *oak* is a kind of *tree*, but a *scrub oak* is a *shrub*, not a *tree.*

A second problem is that the nomenclatural principles have a number of exceptions. For example, the term *dog* is an unanalyzable primary lexeme that labels a category that includes a number of other categories labeled by unanalyzable primary lexemes (*hound*, *setter*, *terrier*), which in turn include categories labeled by secondary lexemes (*basset hound*, *Irish setter*, *Jack Russell terrier*). According to the nomenclatural principles, *dog*

[13] The unique beginner and intermediate categories, if labeled, would violate this nomenclatural principle.

should label a life-form rather than a folk generic. The examples can be multiplied – *cat, horse, cattle* – all have this property. Conversely, the Awajún have a term for epiphytes, *kuish* ("ear"), which refers to a biological range greater than that referred to by some life-form terms (e.g., *shingki* "palm"), but it is not further subdivided. Dougherty (1978) presents a cogent discussion of this and other problems in folk taxonomies. A further difficulty is Gatewood's (1983, 1984) demonstration that, at least among American college students, the names of biological kinds serve as hollow place holders without much in the way of associated knowledge; the students may know the terms *maple*, *oak*, and *elm* as names of kinds of trees, but very often they are unable to identify any actual examples of them.

In addition to the problems mentioned above, which have received a great deal of attention, I believe that there is an even greater problem that is less well recognized. The taxonomies elicited by Berlin et al.'s (1973) approach are generally far too shallow to provide much information about the distinctions that informants readily make among biological kinds. As stated earlier, biological evolution produces very deep binary trees that usually make many feature distinctions at each node of the branching structure. In a similarity judgment task, especially one with design features of the successive pile sort [Boster (1994)], one can elicit judgments of the relative similarities and differences among a collection of specimens that captures the informants' perceptions of this binary tree structure. (A successive pile sort elicits the complete binary tree.) However, if one simply asks which life-forms the specimens belong to (presuming the specimens belong to distinct folk genera), one can only crudely approximate that binary structure. If there were 512 categories at the generic rank in a folk biological classification system, there would be 511 higher order nodes in a complete binary tree, yet there are typically at most ten categories at the life-form rank. What happened to the other 501 superordinate categories? Here again, methods derived from psychology, especially similarity judgment tasks, appear to have greater resolving power than the linguistic methods used by early cognitive anthropologists.

A final problem is the fact that Berlin et al.'s (1973) method assumes that informants agree about classification, but this assumption is patently false – informants often disagree. How do you show that folk and scientists agree on classification without assuming absolute agreement among the folk[14]? Boster, Berlin and O'Neill (1986) provided one solution to this problem by making use of the disagreement itself to show that the folk "confuse" specimens that scientists classify as similar. In this research, bird specimens were placed on long tables and Awajún and Wambis (Aguaruna and Huambisa Jívaro) informants were asked to identify each specimen. There were two groups of bird specimens: passerines, or songbirds, comprising the largest and most recently radiated order of birds (e.g., flycatchers, thrushes, tanagers, and related species) and non-passerine birds, comprising all other orders of birds (e.g., hawks, toucans, and woodpeckers). The naming responses were used to compute an information–theoretic measure of the degree to which pairs of birds were "confused," ranging from 1 if the two bird specimens were given the same name to 0 if there was no overlap in the inventory of names applied to the pair of specimens. This measure of naming overlap was compared with taxonomic

[14] Scientists have more explicit conventions to enforce agreement, although they are not always successful.

distance, a measure of the similarity of the birds in the scientific classification. The authors found that the Awajún and Wambis "confuse" scientifically closely related birds. Furthermore, they found that folk and scientists agree more strongly about non-passerines than about passerines. They concluded that it appears that all three groups use similar classificatory procedures to understand the natural order, but that the clarity of the natural order varies depending on the evolutionary history of the organisms. Biological classification systems are similar only to the extent that the natural order is clear.

Boster (1987) followed up this line of work and completed the shift from linguistic to psychological methods in investigating folk understanding of the biological world. The transition was necessary because the informants were chosen for their biological ignorance and, hence, did not have an elaborate set of named categories for the organisms. The article addressed the following question. It is conceivable that classifiers agree only when they have ample opportunity to study the organisms and have represented their understanding in an explicit nomenclatural system. Will subjects who have no previous knowledge of the birds sort them similarly to the Jívaro and the scientists? In this research, undergraduates with no formal training in zoology nor any familiarity with South American birds were asked to do a successive pile sort of subsets of 15 passerine specimens and 15 non-passerine specimens that had been used as stimuli in Boster (1986). Specimens were chosen such that the passerine and non-passerine subsets had the same underlying scientific taxonomic structure. Boster (1987) found that all groups of subjects (American undergraduates, Awajún, Wambis, and scientists) agreed to a considerable extent in their recognition of patterns of resemblance among the bird specimens and that, as in the earlier research, the correspondence among groups of subjects was stronger for non-passerines than for passerines. Boster (1987) concluded that these diverse human groups appear to perceive the natural order in similar ways and that recognition of that order does not seem to depend on formal training in taxonomy, intimate knowledge of the organisms, or possession of named categories for the specimens.

The next article in this series [Boster and D'Andrade (1989)] was written in response to a question posed by D'Andrade. What if the diverse groups of informants studied in the earlier research were attending to radically different features of the birds, but because the features themselves are intercorrelated, the same structure is discerned no matter what features are chosen? Agreement in classification would not tell us much about the nature of cognition if the informants could not reasonably come up with any other structure. Only if the diverse groups were choosing the same salient features of the specimens to discriminate them, would agreement on classification indicate agreement in the discernment of the organisms.

For example, imagine that the similarities of a gull, a jay, and a cardinal are judged by three informants and that the first informant judges the gull as most different because its webbed feet are different from the perching-bird feet of the cardinal and jay; the second informant judges the gull as most different because its long fish-eating beak is different from the shorter beaks of the cardinal and jay; and the third informant judges the gull as most different because its smooth head is different from the crested heads of the cardinal and jay. All three have judged the gull to be the most different, but each has used a different basis for their judgment. It is because the salient features are correlated that all have come to the same judgment that the two passerines are more similar to each other than either is to the non-passerine.

To address this question, Boster and D'Andrade performed what they called "reverse numerical taxonomy." Numerical taxonomists attempt to derive the best classification of a collection of specimens from a description of each specimen according to a set of features. Boster and D'Andrade reversed this process, using a set of 16 size and color measurements of two subsets of the bird specimens, to discover the relative weighting of the features in the various categorizations of the bird specimens by Awajún, Wambis, undergraduates, and scientists. Boster and D'Andrade found that the diverse groups chose the same features to differentiate each subset of specimens and chose a different set of features to differentiate the non-passerine specimens than they used to differentiate the passerine specimens. The result was interpreted as indicating that members of the different groups use similar perceptual strategies to select those features which will yield the most informative classification of the birds. Boster (1996, p. 283) argued that the human ability to categorize natural kinds demonstrated in these experiments is as much a product of natural selection as the morphological features of the organisms themselves. Summarizing this line of work, he states:

> As the biological world has radiated, the capacity to recognize the order in that radiation has coevolved. The evolution of human cognition to understand the natural world is part of a more general process among living things: mind has evolved to understand nature and leaves its mark as it does so.

7. Towards a science of the stimulus

As a discipline, cognitive anthropology makes its way with a tiny fraction of the personnel of cognitive psychology. But, like other rare endemic species, cognitive anthropologists have their niche. One area of possible contribution is toward a "science of the stimulus," which is often missing from work in cognitive psychology. The reason for its absence is fairly straightforward. Because cognitive psychology is primarily an experimental discipline, it tends to treat any properties of the stimuli as just another variable to control and manipulate and not as something that could be a possible source of explanation in its own right[15]. (The psychologists influenced by Gibson are an important exception to this generalization.)

[15] Inattentiveness to questions about the naturalness of stimuli has a long history in cognitive psychology. In Bruner, Goodnow and Austin's (1956) classic work *A study of thinking*, many of their findings were probably the consequence of using an absolutely artificial stimulus array as their instrument for probing concept formation. The sensitivity of concept formation to the order of presentation of instances that they found was probably a result of the fact that their stimuli had no correlational structure at all. Each of the features of the stimulus array (color, number of instances, shape of figure) varied independently. The same paradigmatic contrasts of features that make for excellent experimental control also make for poor ecological validity. As discussed above, few terms in natural language label combinations of independently varying feature values unless driven by the need for constant reference to the combinations, as is the case with pronominal systems.

In contrast, because cognitive anthropologists ply their trade in natural settings, the properties of the objects of experience have played a more important role in cognitive anthropological explanations of why it is that people understand the world in the way they do. Even those cognitive psychologists who have thought hard and well about the effects that properties of stimuli presented to subjects have on cognitive performance have often not taken their ideas far enough.

For example, Rosch and Mervis (1975) demonstrated that the items in a category that are judged as most typical are those that share the most attributes with other members of the category and the fewest attributes with members of other categories. This is an important insight. However, they did not explain what process generates this particular distribution of attributes. Boster (1988) shows that the typicality ratings of birds elicited by Rosch (1975) from Berkeley undergraduates are more strongly correlated with the number of related species than with the frequency of the birds in the observers' immediate environment or with the frequency of mention of the birds in written materials. Passerines are judged as far more typical than are members of smaller orders of birds. In other words, the pattern in undergraduates' typicality judgments is better explained by data about the phylogeny of the birds being judged than by measures of the subjects' own experience. Correlational structure does not emerge from thin air, but is generated by real world processes – in this case, the process of evolution by means of natural selection.

Although cognitive anthropology emerged as a discipline with a passionate commitment to the possibility of extreme cultural relativism, as the discipline has matured it has documented a number of cultural universals as well as complex patterns of cultural similarities and differences. The methods employed by cognitive anthropologists have also changed over time, as has their understanding of most of the questions addressed. The debates and dead-ends, as well as the cumulative building on each other's accomplishments, show all the hallmarks of a healthy scientific enterprise. In general, the greatest headway has been made when the logic of the methods of data collection and analysis best match the nature of the domain studied.

The domains I have chosen to focus on in this chapter all show either some degree of cross-cultural agreement or a strong pattern in the cross-cultural disagreement. The source of agreement or the pattern in disagreement in each case appears to derive from commonalities in human experience. In the domain of kinship, those commonalities are rooted in the essential facts of life – humans are born, they mate and have children, and they die. Looking back through the generations of our ancestors are pairs of a man and a woman, each of them the son or daughter of other pairs of a man and woman, stretching backwards through time. That genealogical structure, common to all humans, is the basis of cross-cultural universals in kinship terminologies. The regular structure of the genealogical web allows certain relationships to be recognized (e.g., *uncle*, *cìdèèdèè*, *dii?*), just as the regular placement of squares on a chessboard permits the definition of possible moves of the pieces (e.g., "two over and one across" for a knight). However, there is also incredible cultural variation in how kinship is reckoned, so that the biology of reproduction appears to constrain possible kin terminologies no more severely than the use of a particular musical scale constrains the kinds of melodies that can be composed.

Agreement in color classification is more mysterious. Many of us once thought there was a neat neurophysiological explanation of the universals, but it now appears to have been illusory. My hope is that an explanation, when and if it is found, will show that the universals in color classification originate in the interaction between the characteristics of the natural phenomenon itself (the way that the spectrum of light from the sun, filtered through the atmosphere, reflects off of surfaces) and the characteristics of a human visual system that has evolved to make the best use of the information in that radiant signal. I have that hope because that is how it seems to work for human perception and categorization of biological organisms. Biological evolution has produced a marvelous pattern of similarities and differences among species and has also crafted our minds to readily make sense of that pattern.

References

Atkins, J.R. (1974), "GRAFIK: A multipurpose kinship metalanguage", in: Ballonoff, P. A. ed., Genealogical Mathematics (Mouton, Paris) 27–51.

Bellah, R. (1952), Apache Kinship Systems (Harvard University Press, Cambridge, MA).

Berlin, B., D. Breedlove and P. Raven (1966), "Folk taxonomies and biological classification", Science 154:273–275.

Berlin, B., D. Breedlove and P. Raven (1973), "General principles of classification and nomenclature in folk biology", American Anthropologist 74:214–242.

Berlin, B., and P. Kay (1969), Basic Color Terms: Their Universality and Evolution (University of California Press, Berkeley, CA).

Boster, J.S. (1986), "Exchange of varieties and information between Aguaruna manioc cultivators", American Anthropologist 88:429–436.

Boster, J.S. (1987), "Agreement between biological classification systems is not dependent on cultural transmission", American Anthropologist 89:914–919.

Boster, J.S. (1988), "Natural sources of internal category structure: Typicality, familiarity, and similarity of birds", Memory and Cognition 16:258–270.

Boster, J.S. (1994), "The successive pile sort", Cultural Anthropology Methods 6:7–8.

Boster, J.S. (1996), "Human cognition as a product and agent of evolution", in R. Ellen and K. Fukui, eds., Redefining Nature: Ecology, Culture, and Domestication (Berg, Oxford) 269–289.

Boster, J.S., B. Berlin and J.P. O'Neill (1986), "The correspondence of Jivaroan to scientific ornithology", American Anthropologist 88:569–583.

Boster, J.S., and R.G.D'Andrade (1989), "Natural and human sources of cross-cultural agreement in ornithological classification", American Anthropologist 91(1):132–142.

Brown, R. (1965), Social Psychology (Free Press, New York).

Bruner, J.S., J.J. Goodnow and G.A. Austin (1956), "A Study of Thinking", (Wiley, New York).

Campbell, D.T. (1958), "Common fate, similarity, and other indices of the status of aggregates of persons as social entities", Behavioural Sciences 3:14–25.

Conklin, H. (1955), "Hanunóo color categories", Southwestern Journal of Anthropology 11:339–344.

Conklin, H. (1962), "Lexicographical treatment of folk taxonomies", International Journal of American Linguistics 28:119–141. Reprinted in Householder F.W. and Saporta, S. eds., Problems in Lexicography (Indiana Research Center for Anthropology, Folklore and Linguistics, Bloomington, IN) 119–142.

Cook, R.S., P. Kay and T. Regier (this volume), "The world color survey database: History and use", in: H. Cohen and C. Lefebvre, eds., Handbook of Categorization in Cognitive Science(Elsevier, New York).

D'Andrade, R. (1995), The Development of Cognitive Anthropology (Cambridge University Press, Cambridge).

DeValois, R.L., I. Abramov and G.H. Jacobs (1966), "Analysis of the response patterns of LGN cells", Journal of the Optical Society of America 56:966–977.

Dougherty, J. (1978), "Salience and relativity in classification", American Ethnologist 5:66–80.

Frake, C.O. (1962), "The ethnographic study of cognitive systems", in: T. Gladwin and W.C., Sturdevant, eds., Anthropology and Human Behavior (Anthropological Society of Washington, Washington, DC) 72–85.

Frake, C.O. (1964), "Notes on queries in ethnography", American Anthropologist 66:132–145.

Gatewood, J. (1983), "Loose talk: Linguistic competence and recognition ability", American Anthropologist 85:378–387.

Gatewood, J. (1984), "Familiarity, vocabulary size, and recognition ability in four semantic domains", American Ethnologist 11:507–527.

Goodenough, W. (1956), "Componential analysis and the study of meaning", Language 32:195–216.

Goodenough, W. (1957), "Cultural anthropology and linguistics", Georgetown University Monograph Series on Language and Linguistics 3:22–54.

Goodenough, W. (1964), Explorations in Cultural Anthropology, (McGraw-Hill, New York).

Goodenough, W. (1965), "Yankee kinship terminology: A problem in componential analysis", American Anthropologist 67:259–287.

Gould, S.J. (1980), The Panda's Thumb (WW Norton, New York).

Heider, E.R. (1972a), "Probabilities, sampling, and ethnographic method: The case of Dani colour names", Man 7:448–466.

Heider, E.R. (1972b), "Universals in color naming and memory", Journal of Experimental Psychology 93:10–20.

Jameson, K., and D'Andrade (1997), "It's not really red, green, yellow, blue: An inquiry into perceptual color space", in: C.L. Hardin and L. Maffi eds., Color Categories in Thought and Language (Cambridge University Press, Cambridge) 295–319.

Kay, P. (1966), "Comments on "Ethnographic semantics: A preliminary survey" by B.N. Colby", Current Anthropology 7:20–23.

Kay, P., B. Berlin, L. Maffi and W. Merrifield (1997), "Color naming across languages", in: C.L. Hardin and L. Maffi, eds., Color Categories in Thought and Language (Cambridge University Press, Cambridge) 21–56.

Kay, P.,B. Berlin and W. Merrifield, (1991), "Biocultural implications of systems of color naming", Journal of Linguistic Anthropology 1:12–25.

Kay, P., and L. Maffi (1999), "Color appearance and the emergence and evolution of basic color lexicons", American Anthropologist 101:743–760.

Kay, P., and C. McDaniel (1978), "The linguistic significance of the meanings of basic color terms", Language 54:610–646.

Kuhn, T. S. (1962), The Structure of Scientific Revolutions, (University of Chicago Press, Chicago).

Lounsbury, F. G. (1956), "A semantic analysis of the Pawnee kinship usage", Language 32:158–194.

Metzger, D., and G. Williams (1966), "Some procedures and results in the study of native categories: Tzeltal "firewood", American Anthropologist 65:1076–1101.

Miller, G. (1956), "The magical number seven, plus or minus two: Some limits on our capacity for processing information", Psychological Review 63:81–97.

Murdoch, G.P. (1949), Social Structure (MacMillan, New York).

Murdoch, G.P. (1957), "World ethnographic sample", American Anthropologist 59:664–687.

Nerlove, S.B., and A.K. Romney (1967), "Sibling terminology and cross-sex behavior", American Anthropologist 74:1249–1253.

Randall, R.A. (1976), "How tall is a taxonomic tree? Some evidence for dwarfism", American Ethnologist 3:543–553.

Read, D., and C. Behrens (1990), "KAES: An expert system for the algebraic analysis of kinship terminologies", Journal of Quantitative Anthropology 2:353–393.

Romney, A.K. (1965), "Kalmuk Mongol and the classification of lineal kinship terminologies", American Anthropologist 67:127–141.

Romney, A.K., and R.G. D'Andrade (1964), "Cognitive aspects of English kin terms", American Anthropologist 68(3)2:146–170.

Rosch, E. (1975), "Cognitive representations of semantic categories", Journal of Experimental Psychology: General 104:192–253.

Rosch, E., and C.B. Mervis, (1975), "Family resemblances: Studies in the internal structure of categories", Cognitive Psychology 7:573–605.

Schneider, D.M. (1965), "American kin terms and terms for kinsmen: A critique of Goodenough's componential analysis of Yankee kinship terminology", American Anthropologist 67:288–308.

Shweder, R.A., and E.J. Bourne (1984), "Does the concept of the person vary cross-culturally?" in: R.A. Shweder and R.A. LeVine, eds., Culture Theory: Essays on Mind, Self, and Emotion (Cambridge University Press, Cambridge) 158–199.

Spencer, H. (1897), Principles of Sociology, 3rd edition (Appleton,New York) (1st edition, 1876.)

Tyler, S. ed (1969), Cognitive Anthropology, (Holt, Rinehart, & Winston, New York).

Wallace, A. (1964), "On just being complicated enough", Proceedings of the National Academy of Science 17:458–461.

Wallace, A., and J. Atkins, (1960), "The meaning of kin terms", American Anthropologist 62:58–80.

Woolford, E. (1984), "Universals and rule options in kinship terminology: A synthesis of three formal approaches", American Ethnologist 11:771–790.