Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 4699–4705, May 1996
Anthropology

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 25, 1995

Culture as shared cognitive representations

(culture consensus/correspondence analysis/kinship terms/multidimensional scaling/semantic structure)

A. KIMBALL ROMNEY*, JOHN P. BOYD, CARMELLA C. MOORE, WILLIAM H. BATCHELDER, AND TIMOTHY J. BRAZILL

School of Social Sciences, University of California, Irvine, CA 92717

Contributed by A. Kimball Romney, February 15, 1996

ABSTRACT **Culture consists of shared cognitive representations in the minds of individuals. This paper investigates the extent to which English speakers share the "same" semantic structure of English kinship terms. The semantic structure is defined as the arrangement of the terms relative to each other as represented in a metric space in which items judged more similar are placed closer to each other than items judged as less similar. The cognitive representation of the semantic structure, residing in the mind of an individual, is measured by judged similarity tasks involving comparisons among terms. Using six independent measurements, from each of 122 individuals, correspondence analysis represents the data in a common multidimensional spatial representation. Judged by a variety of statistical procedures, the individuals in our sample share virtually identical cognitive representations of the semantic structure of kinship terms. This model of culture accounts for 70–90% of the total variability in these data. We argue that our findings on kinship should generalize to all semantic domains—e.g., animals, emotions, etc. The investigation of semantic domains is important because they may reside in localized functional units in the brain, because they relate to a variety of cognitive processes, and because they have the potential to provide methods for diagnosing individual breakdowns in the structure of cognitive representations typical of such ailments as Alzheimer disease.**

Recent methodological advances make possible precise comparisons among cognitive representations in the minds of different individuals (1–3). We can now measure with known accuracy the extent to which "pictures" or cognitive representations in the mind of one person correspond to those in the mind of another. Not only can we measure the extent to which a large number of individuals "share" the same picture, but we can make multiple measures of the picture in the mind of a single individual. The picture in the mind of a single individual may be thought of as a representation of the structure of some semantic domain such as kinship terms.

A semantic domain may be defined as an organized set of words, all on the same level of contrast, that refer to a single conceptual category, such as fish, furniture, or vehicles (4). (Note that a semantic domain does not include the term for the superordinate category.) The structure of a semantic domain is defined as the arrangement of the terms relative to each other as represented in some metric system, such as Euclidean space, and described as a set of interpoint distances reflecting the dissimilarity between them. In this space, items that are judged more similar are closer to each other than items that are judged less similar.

Semantic domains may be localized functional units in the brain. Neuropsychological studies have shown that aphasic patients sometimes have selective impairment of specific semantic categories such as flowers, vegetables, or animals (5–9). The concept of semantic structure also appears in investigations of Alzheimer and Huntington diseases, where there is a breakdown of semantic structure as well as a deterioration of its accessibility (10, 11).

The structure of a semantic domain may be interpreted as a cognitive representation derived from judged-similarity tasks. This paper demonstrates how such measurements are made and discusses the reliability of these scaled representations. An important assumption is that the resulting spatial cognitive representation is, in some sense, isomorphic with what is in the mind of the subject (12). If this assumption is correct, then the representation should predict a variety of cognitive processes. Distances in such models have been shown to predict categorical judgment time (13–15), completion of analogies (14, 16), the strength of semantic clustering in memory (4), and response times in solving triadic comparison problems (17, 18). These findings illustrate the usefulness of these representations in cognitive science and the potential for their wider application in anthropology and other behavioral sciences.

We note that the evolution of language and cultural knowledge, including all of scientific knowledge, is totally dependent upon the sharing of linguistic meanings. The learning of these shared meanings by individuals as members of human groups is a highly selected-for human survival skill transmitted from generation to generation. This view is consistent with the new developments in the field of evolutionary psychology (19).

It is assumed, then, that each individual has an internal cognitive representation of the semantic structure in which the meaning of a term is defined by its location relative to all the other terms. Culture consists of shared cognitive representations of this structure. Our aim here is to measure the extent to which a number of individuals share cognitive representations and to identify any systematic differences of representations among subgroups of subjects.

Kinship was chosen as the semantic domain for analysis because it is purely cultural in content. Kin terms are abstract concepts; they cannot be characterized in terms of obvious external physical characteristics or as occupying a single visible location. Kin terms have no "concrete" referents, such as size or color, in the way an animal such as "dog" or "cat" has. Different societies have very different ways of categorizing relatives; the English system is only one of many. For example, in English the categories of "mother" and "mother's sister" are distinguished by separate terms (i.e., mother and aunt), whereas in many societies "mother's sister" is called by the same term as "mother."

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: BIBD, balanced incomplete block design.
*To whom reprint requests should be addressed.

Kinship is one of the domains studied earliest in anthropology and one of the more theoretically and technically developed domains. The first scaling of kinship terms using judged similarity was performed by Romney and D'Andrade (20), who related the semantic structure to componential analysis for the eight basic male kin terms. They predicted that the more components any two terms had in common, the greater the similarity of response to these terms as judged in a triads test. They assumed that the components of a term constituted the meaning of that term for an individual; hence, the more components in common, the more similar the meaning between terms. Romney and D'Andrade (20) demonstrated that a single aggregated cognitive representation, based on judged similarities collected with the triads task, corresponded closely with only one of the alternative models posited at that time. Two studies in the early 1970s extended the scaling to all 15 basic terms (21, 22). A number of subsequent studies collected similarity data and presented spatial models in basic agreement with Romney and D'Andrade's findings (23–26). The current study builds on this tradition and extends its scope.

METHODS

Sample. Our subjects were 122 University of California, Irvine, undergraduates from a variety of ethnic and linguistic backgrounds. We collected data on gender as well as data on languages that were learned in childhood. On the basis of language we classified subjects into three categories: English only, English first, and English second. The first category subjects were monolingual in English as children, the second category subjects learned two or more languages as children with English as the first language learned, and the third category subjects learned English as a second or subsequent language. The most frequently occurring languages other than English were Vietnamese (n = 17), various Chinese dialects (n = 15), Korean (n = 13), Spanish (n = 10), Tagalog (n = 5), and a variety of others. Table 1 shows the distribution of subjects by gender and language category.

Triads Task and Design. Data on judged similarities were collected using the triads test. The 15 English kinship terms are grandfather (GrFa), grandson (GrSo), father (Fa), son (So), brother (Br), uncle (Un), nephew (Ne), cousin (Co), grandmother (GrMo), granddaughter (GrDa), mother (Mo), daughter (Da), sister (Si), aunt (Au), and niece (Ni). We collected judged similarity data with questionnaires consisting of 210 triads of terms. The task was to pick, for each triad, the term that seemed most different in meaning from the other two (27). We used a balanced incomplete block design (BIBD) derived by one of the authors (J.P.B.) that produced six equivalent but disjoint (i.e., no triad is repeated) sets of 35 triads. Each of the six sets consisted of a lambda-one BIBD (see ref. 27, pp. 49–53), in which every pair of terms occurs exactly once. This provided information that allowed us to obtain six independent scaling results for each subject. The order within and between triads was individually randomized across all six sets of lambda-one BIBDs for administration of the questionnaire.

The data for each subject were scored in a series of six 15 × 15 symmetric binary matrices with the rows and columns labeled by the kin terms in the order indicated earlier. Each matrix contained the raw similarity data from a single lambda-one BIBD set as follows. For each triad (of the set of 35 triads in a specific BIBD), a point was entered in the matrix representing the pair remaining after the subject picked the one most different from the other two. If, for example, in the triad of MOTHER, FATHER, COUSIN, the subject had circled COUSIN as most different, then a point of similarity was entered for the pair MOTHER–FATHER in the cell representing the intersection of MOTHER and FATHER.

Table 1. Subjects classified by gender and language category

	Female	Male	Total
English only	22	21	43
English first	6	4	10
English second	36	33	69
Total	64	58	122

Correspondence Analysis of Stacked Similarity Matrices. To represent all subjects in the same space, we stacked the data from all subjects and tasks into a single matrix and performed a correspondence analysis on that matrix. Because correspondence analysis assumes similarity data, a "one" was placed on the diagonal of each matrix on the assumption that each item was maximally similar to itself (28). The stacked matrix contained (122 × 6) 732 matrices. Because there were 15 terms in each matrix, it had 10,980 rows (15 × 732) and 15 columns (one for each term). Analysis of the stacked matrix produced a multidimensional scaling representation of the data. The software utilized was SAS (29). The analysis is standard and can be found in any conventional treatment of correspondence analysis (28, 30–32).

Weighted optimal scores were obtained by weighting the row scores by the square root of the singular values. These scores are used in the analysis of variance reported below and will be referred to as unstandardized. The use of these scores for plotting purposes would have meant comparing individual pictures of different sizes because of differences in variances in the responses given by different subjects or in different tasks (1, 2). These differences in scale (size) were viewed as artifacts of individual differences of subjects or of task differences in filling out the triads and without any substantive interpretation. Because we wanted to compare tasks within individual subjects as well as among subjects, we needed to correct for these differences among subjects and tasks.

Common practice in correspondence analysis does not prescribe a standard transformation for bringing scores within each subject to scale. Consequently, we devised a method of transforming row scores within each subject that would correct the problem of scale by standardizing the scores for each subject to zero mean and variance equal to the square root of the singular values. This procedure had previously been utilized in a study of biases in social perception (1) and in a study of the scaling of semantic domains (2). These standardized scores were used in plotting all figures and in computing goodness-of-fit measures.

Goodness-of-Fit Measures of Scaling. A measure of the "resolving power" of the final scaling model obtained by the correspondence analysis was needed to ascertain how well the data were fit by the description. In previous work we had found that a simple measure of the proportion reduction of error, η^2, gave satisfactory results (2, 3). The measure was η^2 obtained by performing a one-way analysis of variance, using "kin term" as the category grouping, on the standardized row scores for each of the dimensions and is simply the sum of squares explained by the category of kin terms divided by the total sum of squares.

The motivation for η^2 is provided by the following observations. If every person had exactly the same cognitive representation of the structure of the semantic domain of kinship and if there were no measurement errors whatsoever, then all individuals would place each term in the exact same position. There would be no variability among subjects, and knowledge of which term was being considered would determine the location completely, giving an η^2 of 1.00. Conversely, no agreement among subjects would allow no knowledge of location to be obtained by knowing which term was being considered and would produce an η^2 approaching zero.

Cultural Consensus and Individual Competence. To obtain another perspective on cultural knowledge, we asked how much knowledge of the cultural content of the triads questionnaire was possessed by each individual and whether some subgroups of individuals had more knowledge than others. Cultural consensus analysis consists of a family of formally derived mathematical models that simultaneously provide an estimate of the cultural competence or knowledge of each informant and an estimate of the correct answer to each question asked. Formal consensus process models have been derived from basic axioms for some formats of data collection (33–35) such as dichotomous and multiple choice. A less formal, data-level model has been derived for rank order data (36). Because triad data are unlikely to satisfy one of the three basic axioms of the formal process model, namely, local independence, we developed a general data-level model for the estimation of cultural consensus parameters for such data.

The consensus model provides a way to utilize much of the accumulated knowledge of traditional psychometric test theory without knowing the "correct" answers in advance. Whereas traditional test theory begins with "performance" data (i.e., items coded as "correct" or "incorrect"), consensus theory begins with "response" data (items coded as given by the informant; for example, "true" or "false," with no assumptions about whether the informant is correct or incorrect). The potential implications of this fact for the behavioral sciences may be important. It means that we are now in a position to measure the knowledge of subjects where we do not know the answers to the questions we ask and to do so with a degree of accuracy comparable to that obtained with traditional test theory.

The model has been subjected to extensive testing through simulation (37, 38) and Monte Carlo methods. It has been applied to folk medical beliefs (39), judgment of personality traits in a college sorority (40), semiotic studies of alphabetic systems (41), occupational prestige (18), causes of death (36), hot-cold illness concepts (42–44), child abuse (45), and social network data (46).

The formal process model for cultural consensus was designed for very small numbers of informants (e.g., half a dozen or so). It works at this level for three reasons: (*i*) the theory is precise, with very strong assumptions; (*ii*) we work in a range of high-concordance codes where consensus is high; and (*iii*) we are only trying to find one "correct" answer for a question rather than differentiating questions on a continuous scale. In normal applications of consensus analysis there should be good reasons for assuming that the axioms are not violated in a serious way by the data. With small samples and when the axioms appear reasonable, the competence of the informants is estimated directly from the data. The answer key is then estimated by weighting informants' contributions to the answer key proportional to their competence. In this way the more "knowledgeable" subjects contribute more to the inference of the answer key than do less "knowledgeable" informants. In small samples this procedure is crucial.

In the case of triads, where the answer for some triads may be strongly related to the answer for other triads, one of the axioms of consensus analysis, that of local independence, is violated. This means that we cannot use procedures that depend on the independence assumption without biasing our estimates to a significant extent. Our solution to this problem applies in general to the triads format of collecting judged similarity data in the case where we have large numbers of subjects (i.e., more than 30).

In our data-level model for estimating the competence of each individual, the answer key is estimated before the estimate of individual competence on the assumption that each subject contributes an equal amount of information to the answer to each question. Even though this assumption is unlikely to be true, with large numbers of subjects the modal response converges toward the correct answer as the number of subjects increases. This means that, as D'Andrade (47) has suggested, the modal response can be used for each triad to estimate the answer key. Given the answer key the remainder of the multiple-choice model is specified by Romney *et al.* (ref. 33, p. 319) as follows:

> In case the correct answer key is known it is easy to simply count the number of correct responses and divide by M, the number of questions, to obtain the proportion of correct responses T_i for informant i. In order to obtain an estimate of D_i [the cultural competence of subject i] we use the empirically observed T_i . . . [as the probability of any item being correctly answered by an informant] . . . and solve for D_i to obtain
>
> $$\hat{D}_i = (LT_i - 1)/(L - 1), \quad (4)$$
>
> where [to adapt formula (4) to a triads test, let $L = 3$] the hat over the D_i is the usual convention to indicate that it is an estimate of the underlying competency D_i. All Eq. 4 does is to adjust the proportion correct for guessing, and this is used routinely in aptitude testing by ETS and other agencies.

A bonus is provided by the design of the present study, which makes possible the calculation of the reliability of the measure of cultural competence. Since we have six estimates of each individual's competence (one for each of the six lambda-one BIBD sets of 35 triads), we can apply the Spearman–Brown (48, 49) reliability formula (also called Cronbach's α), given by Nunnally (50, p. 193) as

$$r_{kk} = \frac{k\bar{r}_{ij}}{1 + (k - 1)\,\bar{r}_{ij}},$$

where $\bar{r}_{ij}$ is the mean correlation among tests, k is the number of items or tests, and r_{kk} is the reliability coefficient for a k-item test determined from the intercorrelations of items on the test. The equation holds regardless of the size of the units that are added. In our case the units are the six separate lambda-one triad tests and r_{kk} refers to the reliability of the overall competence score derived from the total 210 items. The square root of the reliability score estimates the correlation between the "test" and the "truth."

RESULTS

Graphical Display of Scaling. Visualizable spatial models capture the essential structure in the data and display in comprehensible form both the stability and the variability in the data as captured in the scaling results. These spatial representations provide a background for a deeper understanding of the statistical results. The correspondence analysis resulted in 10,980 row optimal scores and three dimensions were retained, with singular values of 0.644, 0.483, 0.388, respectively, accounting for 48% of the variance. Illustrative plots are presented in only two dimensions. There are 732 row optimal scores for each kin term–six task scores for each of the 122 subjects. The two-dimensional plot of all 10,980 scores forms an incomprehensible cloud of points, but by judicious choice of subsets of points to plot we can summarize and contrast any desired aspects of the data.

Perhaps the single most important visual summary of the data bearing on the theory of culture as shared cognitive representations is obtained by an aggregate global view of how all subjects placed each kin term as shown in Fig. 1. The ellipses represent 95% confidence limits on the mean scores for each term. That mean is considered the cultural definition of the term and appears as the solid square in the center of the corresponding labeled ellipse. The confidence ellipses are estimated from all 732 spatial locations (six placements for each of 122 subjects) of each term under a bivariate normal assumption, and they give a visual idea of the degree of resolution of the methods. Note that terms that are close to each other in the first two dimensions may be differentiated in higher dimensional representations.

There must clearly be a very large area of agreement among individuals to produce such sharply limited areas of estimated location of the terms relative to each other. It should be noted

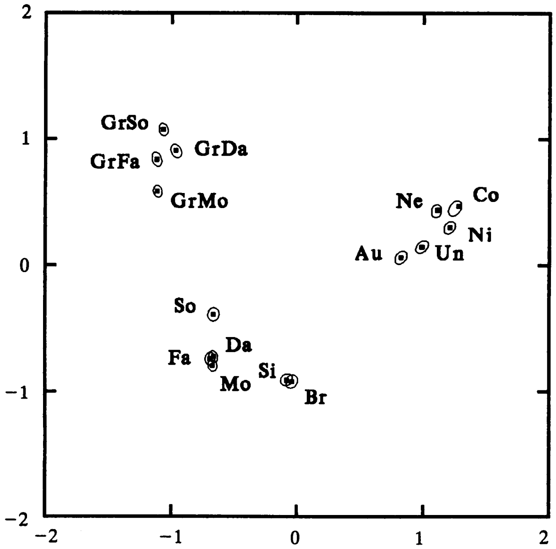

FIG. 1. Semantic structure of kinship terms with cultural position defined as the mean of all subjects (■) together with 95% confidence ellipses of that mean (computed from 122 subjects with 6 tasks each).

that the area of the confidence ellipses is a function of both variability in the measures and the number of cases involved. The fact that each ellipse is based on 732 scores accounts in large part for how small they are. In subsequent plots we will see larger ellipses as a function of both reduced number and greater variability among individuals.

Our methods also lend themselves to the comparison of configurations among subgroups. For example, the comparison in Fig. 2 between females and males shows visually that there is virtually no difference between females and males because the ellipses for the two groups touch or overlap for every kin term. The two configurations are virtually identical to the global configuration in Fig. 1.

Fig. 3 compares monolingual English speakers with those who learned English as a second language. Although the ellipses occupy similar configurations in general, those who learned English as a second language have more variability among themselves, as indicated by larger ellipses for every

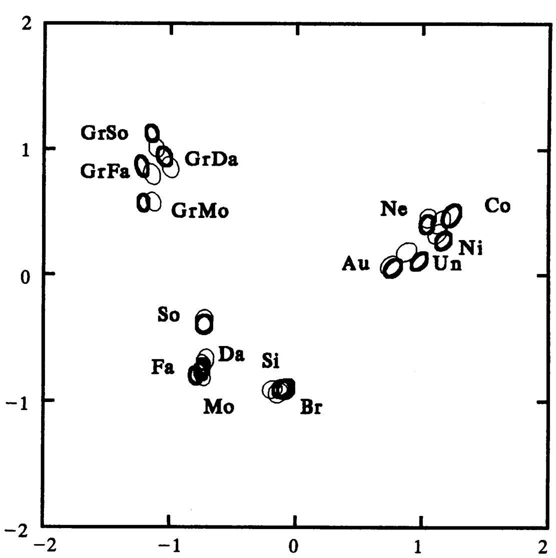

FIG. 2. Semantic structure of kinship terms represented as 95% confidence ellipses of the mean position of each term comparing 64 female subjects (bold outlined ellipses) with 58 male subjects (light outlined ellipses).

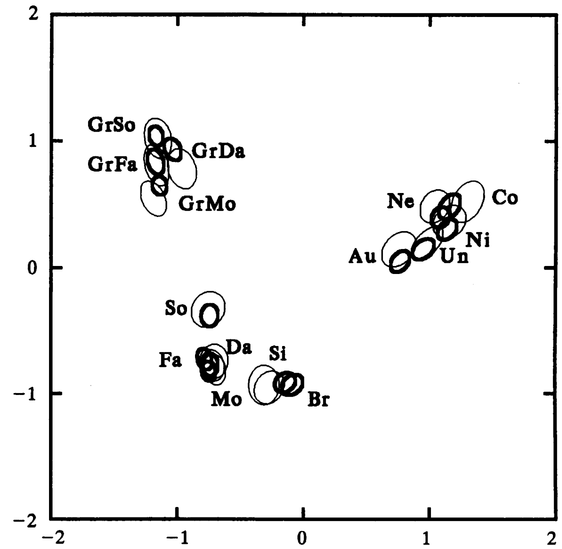

FIG. 3. Semantic structure of kinship terms represented as 95% confidence ellipses of the mean position of each term comparing 43 monolingual English speakers (bold outlined ellipses) with 69 subjects with English as a second language (light outlined ellipses).

term. This is not a function of size of sample, as there are more subjects with English as a second language (n = 69) than monolingual English speakers (n = 43).

That our methods do not depend on large numbers of subjects is apparent from the contrast between small samples from the 122 original subjects. On the basis of our estimates of cultural competence (to be reported later) we contrasted the 10 most competent ($D_i > 0.725$) subjects with the 10 least competent ($D_i < 0.350$) subjects in Fig. 4. We found that the more competent subjects produced results very similar to the global cultural configuration illustrated in Fig. 1 with, of course, larger ellipses because of small sample size. The general picture of the 10 least competent subjects is still in relatively good agreement with the overall configuration. The ellipses are larger for this group, indicating much more variability among the less competent subjects, as would be expected. An inherent characteristic of the method is that as agreement declines among subjects the ellipses get larger and at the same time drift to the center of the figure.

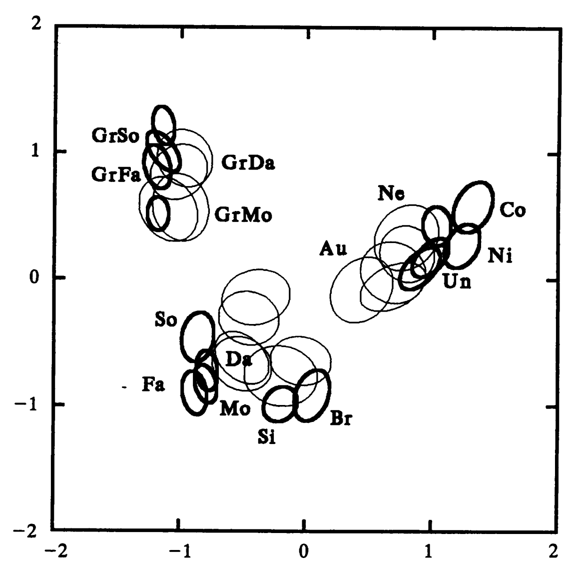

FIG. 4. Semantic structure of kinship terms represented as 95% confidence ellipses of the mean position of each term comparing the 10 most competent subjects (bold outlined ellipses) with the 10 least competent subjects (light outlined ellipses).

In the case of no agreement, all the ellipses would be large and centered on the midpoint of the diagram.

In many kinds of research it may be desirable to look at the patterns of individual people. The methods we have presented facilitate viewing individuals and comparing them to the cultural pattern. Clearly, we cannot present a picture for each of the 122 subjects. Instead, we have selected one subject from the upper quartile of competence. Fig. 5 shows the fit between the culture and that individual's cognitive representation. The placement of terms in this figure is represented as the mean of that individual's performance on the six lambda-one tasks.

We have presented only a small sampling of possible plots to illustrate the kinds of information they contain. Others might include a plot showing the location of the above subject for each of the six tasks, a plot of the location of all 122 individuals as a vector from the cultural definition of the terms, and a plot of each task with ellipses for the terms.

Descriptive Analysis of Variance. A mixed model of analysis of variance for the categorical variables of language and gender and for repeated measures of task and terms for the domain of kinship terms produced the results shown in Table 2. We present a separate analysis of the unstandardized row scores for each of the first three dimensions. Each analysis is based upon the 10,980 row scores.

What results might be expected from such an analysis? The received view in cultural anthropology today would anticipate that such an analysis would show that gender and language made a large difference, that males and females had a somewhat different picture in their minds about the similarity pattern among kinship terms, and that individuals who learned English as a second language had different definitions of similarity than monolingual English speakers. According to the received view, complex interactions among task and term and other variables might have been predicted. Our own expectation was that term would have the largest effect, because distinctions among kin terms constitutes the essence of cultural sharing. As we have noted, if all individuals agreed perfectly on the pattern of judged similarities with no measurement error whatsoever, then all variance would be accounted for by term. In practice we would expect some relatively small amount of variance to be associated with task, since each task contains different triads of words to be judged in terms of similarity; therefore tasks might differ somewhat from one another. Thus task differences would reflect components of measurement error and should be small compared with differences in term. Task-by-term interaction could occur and would indicate that the different tasks affected different kinship terms differentially. Because we believe that culture is highly shared, we did not expect any major differences associated with language or gender.

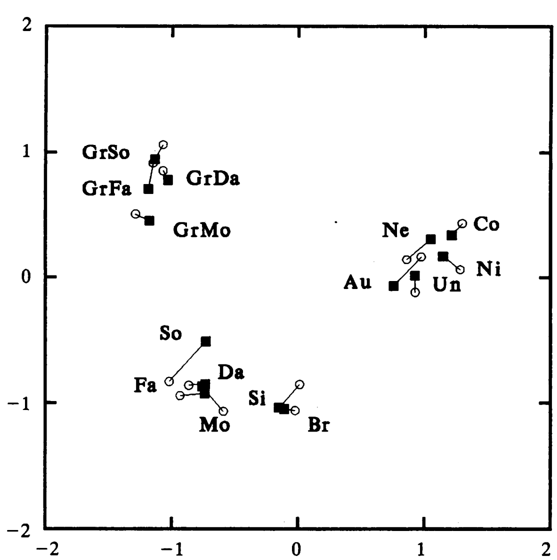

FIG. 5. Semantic structure of kinship terms comparing the mean (of six tasks) position of a single subject (○) with the cultural position (■) for each term.

Table 2. Results for mixed model ANOVA on unstandardized scores for dimensions one to three for categorical variables of language and gender and for repeated measures of task and terms (interactions treated as error)

Source	SS	df	MS	*F*	*P*
Dimension one					
Between subjects					
Language	0.529	2	0.265	7.713	0.001
Gender	0.001	1	0.001	0.035	0.852
Error	3.982	118	0.034		
Within subjects					
Term	2031.897	14	145.136	573.660	0.000
Error	428.618	1694	0.253		
Task	0.654	5	0.131	12.516	0.000
Error	6.292	605	0.010		
Term·Task	106.959	70	1.528	20.433	0.000
Error	633.327	8470	0.075		
Dimension two					
Between subjects					
Language	0.108	2	0.054	1.023	0.363
Gender	0.000	1	0.000	0.000	0.998
Error	6.153	118	0.053		
Within subjects					
Term	858.732	14	61.338	312.950	0.000
Error	332.284	1694	0.196		
Task	0.619	5	0.124	8.296	0.000
Error	8.834	605	0.015		
Term·Task	113.577	70	1.623	22.384	0.000
Error	609.289	8470	0.072		
Dimension three					
Between subjects					
Language	0.053	2	0.027	0.567	0.569
Gender	0.070	1	0.070	1.495	0.224
Error	5.556	118	0.047		
Within subjects					
Term	468.255	14	33.447	200.566	0.000
Error	283.167	1694	0.167		
Task	1.935	5	0.387	16.466	0.000
Error	13.953	605	0.023		
Term·Task	349.423	70	4.992	60.745	0.000
Error	696.571	8470	0.082		

Table 2 contains the empirical results of an analysis of variance that support several general conclusions. First, gender makes no discernible contribution to variance. This rules out a female–male difference in cognitive representations for judged similarity among kinship terms. Second, language background makes a difference only in the first dimension. Even though the difference is statistically highly significant, it is in fact very small compared to the effects of term, task, and term-by-task interaction. This suggests that the semantic structure of kinship does not depend on when English is learned, because those who learned it as a second language are barely different from those who learned it first or as their only language. Third, term accounts for a very large part of the total variance—almost two-thirds of the total sums of squares of the first dimension. This result is not inconsistent with the idea that all individuals in the study share the "same" representation of the semantic structure. Finally, task has a statistically significant, although in absolute terms a fairly small, effect. The term-by-task interaction is larger than the task effect, although it is only a small fraction as large as the term effect, especially for the first dimension.

The consistency of the pattern of results across the various dimensions, except for language, is noteworthy. The effects are certainly most dramatic for the first dimension and become considerably diluted by the third dimension. However, the importance of term effects is clear.

Model Fit Measures. The results of calculating η^2 as a proportion-reduction-error measure are given in Table 3, which reports results for each task taken individually as a lambda-one design and for the pooled results of treating the whole task as a single lambda-six design. Results for gender for the lambda-six design are also included. In all cases the analysis is based on 1830 scores (122 subjects by 15 kin terms) using a simple one-way analysis of variance with "term" as the category. The measure is calculated separately for each of the first three dimensions.

The lambda-one tasks are inherently less reliable than the single lambda-six task and, as expected, have much lower η^2s. The first dimension η^2s of these tasks average about 0.80 and are sufficient to provide good cognitive representations and good results when related to other cognitive processes. The second dimension η^2s average about 0.66, which is borderline, while the third dimension averages about 0.48, which is lower than one would desire. We have no explanation for the outlying value of 0.86 for the third dimension for the first task. These scores can be compared with those obtained for a lambda-one design for 21 animals reported by Romney *et al.* (2). For two separate sets of subjects scores were between 0.72 and 0.76 for the first dimension, between 0.67 and 0.74 for the second dimension, and between 0.46 and 0.56 for the third dimension.

Even better results would be expected with designs that are based on more information and hence have higher reliability and, presumably, higher validity. The lambda-six design η^2s for the first three dimensions of 0.94, 0.88, and 0.73 are quite remarkable. These figures are based on a total of 210 triads (while the lambda-one designs had only 35 triads for the 15 kinship terms). They can be compared with a paired-comparison task involving 210 pairs that were judged on a seven-point scale of similarity among 21 animals reported in the Romney *et al.* (2) study. Figures for two samples were 0.93 and 0.95 for the first dimension, 0.90 and 0.93 for the second dimension, and 0.85 and 0.88 for the third dimension. We can see how good an η^2 of 0.94 is when we reflect that this means that knowing which term the score is measuring accounts for 94% of the variance in the scores for the first dimension. This leaves a total of 6% to cover all individual differences among subjects as well as measurement and sampling error from all sources. Table 3 also shows η^2 for females and males separately; female subjects have a slightly larger η^2 than males.

Individual Competence. The competence of each subject was calculated across the 210 triads using the modal response as the answer key. The mean competence was 0.58 (SD = 0.14) after correction for guessing (the actual mean proportion correct was 0.72). There were no significant differences related to language. There was, however, a gender difference, with females having higher competence than males. Females had a mean competence of 0.62 (SD = 0.11) compared with a mean of 0.53 (SD = 0.15) for males, a highly significant difference (t = 3.78; $P < 0.001$).

We also calculated the competence for each individual on each of the six separate lambda-one designs. The mean correlation among the six tasks was 0.60. From this we calculated a Spearman–Brown reliability coefficient (or Cronbach's α) of 0.90. This indicates that our estimates are highly reliable and that we would expect other researchers to be able to replicate our results. We note that the square root of 0.90 is about 0.95, which would be an estimate of the correlation of our overall triads task of 210 items with the "truth."

Table 3. η^2 for three dimensions of the six lambda-one BIBD tasks and for a single lambda-six BIBD, consolidated and by gender, computed on the standardized scores

	Dimension		
	1	2	3
Lambda-one BIBD task			
1	0.84	0.70	0.86
2	0.81	0.71	0.46
3	0.80	0.66	0.45
4	0.78	0.65	0.45
5	0.82	0.73	0.34
6	0.75	0.53	0.32
Lambda-six BIBD			
taken as single task			
Total sample	0.94	0.88	0.73
Female	0.95	0.91	0.77
Male	0.92	0.86	0.70

DISCUSSION

The major aim of this research was to provide a quantitative model of culture as shared cognitive representations of the structure of semantic domains. We recognize that shared meanings in semantic domains are only a small part of culture. They do, however, provide a simple, ideal natural unit as a model system. The specific outcome of this research is to support the view that typical English speakers share a cognitive representation of the judged similarity among the 15 basic kinship terms.

An interesting and important corollary of our main findings is that social scientists can safely apply cultural definitions of semantic structures in predicting individual cognitive behaviors. In fact, the cultural definition is a better estimate of what is in the mind of the subject than an estimate of a cognitive representation based on the subject's own responses. This is because of the vastly increased reliability of aggregate measures compared with single measures that is apparent from an examination of our figures. This observation simply validates what has been a *de facto* practice in psychology for decades, most multidimensional representations having been, up until now, single pictures based on aggregate data.

The results also imply that the structure of semantic domains is routinely learned with the language. The language category had little effect on cultural competence and only minimal effect in the analysis of variance. There was somewhat more variance among those who had learned English as a second language than among monolingual English speakers, but if the structure of semantic domains were really very difficult for human beings to acquire we would have expected a greater effect of this variable.

We should stress that our conclusions are based on a single study using very simple criteria for measuring the degree of sharing of cognitive representations among individuals. Before accepting these results and conclusions with any great degree of confidence, further research needs to be carried out in two critical directions. First, there should be studies that include other domains such as animals, colors, and emotions. Second, a process model, of the sort exemplified in some of our earlier work (4, 34), should be derived to model parameters that would measure the exact ratio of variance accounted for by individual differences relative to overall error variance. Such a model would augment the scaling approach we have used in three important ways: (*i*) it would provide a processing account of the data in terms of interpretable parameters that reflect subject characteristics, item and task characteristics common to all subjects, and measurement error; (*ii*) it would allow data to be simulated and scaled with various parameter settings to assist in validating the interpretation of the empirical representations; and (*iii*) in favorable cases, it could be subjected to goodness-of-fit and hypothesis tests. A possible disadvantage of the modeling approach is that it is necessary to make very simple assumptions that are specific to a particular data format.

In the meantime, this research has reported methods for scaling into a single representation similarity judgments based on data from many subjects each performing multiple tasks. In addition to shedding light on the concept of culture as shared cognitive representations of the individuals in the culture, the methods can serve other major purposes. First, they permit the examination of very large data sets that otherwise would not be accessible in a single coherent view. Second, they allow the description and testing of comparisons among individuals and subgroups in a completely flexible manner. Third, they provide an optimally aggregated cultural representation that can be used to predict cognitive behaviors related to cognitive structure. Finally, they may be useful for diagnosing individual breakdowns in the structure of cognitive representations in situations such as Alzheimer and Huntington diseases.

This research was supported by National Science Foundation Grant SES-9210009 to A.K.R. and W.H.B.

1. Kumbasar, E., Romney, A. K. & Batchelder, W. H. (1994) *Am. J. Sociol.* **100,** 477–505.
2. Romney, A. K., Batchelder, W. H. & Brazill, T. J. (1995) in *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday*, eds. Luce, R. D., D'Zmura, M., Hoffman, D., Iverson, G. J. & Romney, A. K. (Lawrence Erlbaum, Mahwah, NJ), pp. 267–294.
3. Brazill, T. J., Romney, A. K. & Batchelder, W. H. (1995) *J. Quant. Anthropol.* **5,** 359–374.
4. Romney, A. K., Brewer, D. D. & Batchelder, W. H. (1993) *Psychol. Sci.* **4,** 28–34.
5. Goodglass, H., Wingfield, A., Hyde, M. R. & Theurkauf, J. C. (1986) *Cortex* **22,** 87–102.
6. Sartori, G. & Job, R. (1988) *Cognit. Neuropsychol.* **5,** 105–132.
7. Silveri, M. C. & Gainotti, G. (1988) *Cognit. Neuropsychol.* **5,** 677–709.
8. Warrington, E. K. & McCarthy, R. A. (1983) *Brain* **106,** 859–878.
9. Warrington, E. K. & Shallice, T. (1984) *Brain* **107,** 829–854.
10. Chan, A. S., Butters, N., Paulsen, J. S., Salmon, D. P., Swenson, M. R. & Maloney, L. T. (1993) *J. Cognit. Neurosci.* **5,** 254–261.
11. Chan, A. S., Butters, N., Salmon, D. P. & McGuire, K. A. (1993) *Psychol. Aging* **8,** 411–419.
12. Shepard, R. N. & Chipman, S. (1970) *Cognit. Psychol.* **1,** 1–17.
13. Caramazza, A., Hersh, H. & Torgerson, W. S. (1976) *J. Verb. Learn. Verb. Behav.* **15,** 103–117.
14. Rips, L. J., Shoben, E. J. & Smith, E. E. (1973) *J. Verb. Learn. Verb. Behav.* **12,** 1–20.
15. Shoben, E. J. (1976) *J. Verb. Learn. Verb. Behav.* **15,** 365–379.
16. Rumelhart, D. L. & Abrahamson, A. A. (1973) *Cognit. Psychol.* **5,** 1–28.
17. Hutchison, J. W. & Lockhead, G. R. (1977) *J. Exp. Psychol. Hum. Learn. Mem.* **3,** 660–678.
18. Romney, A. K. (1989) *J. Quant. Anthropol.* **1,** 153–223.
19. Barkow, J. H., Cosmides, L. & Tooby, J. (1992) *The Adapted Mind* (Oxford Univ. Press, New York).
20. Romney, A. K. & D'Andrade, R. G. (1964) *Am. Anthropol.* **66,** 146–170.
21. Nerlove, S. B. & Burton, M. L. (1972) *Am. Anthropol.* **74,** 1249–1253.
22. Fillenbaum, S. & Rapoport, A. (1971) *Structures in the Subjective Lexicon* (Academic, New York).
23. Wexler, K. N. & Romney, A. K. (1972) in *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, eds. Romney, A. K., Shepard, R. N. & Nerlove, S. B. (Seminar, New York), Vol. 2, pp. 73–92.
24. Rosenberg, S. & Kim, M. P. (1975) *Multivar. Behav. Res.* **10,** 489–502.
25. Arabie, P., Carroll, J. D. & DeSarbo, W. S. (1987) *Three-Way Scaling and Clustering* (Sage, Newbury Park, CA).
26. Romney, A. K., Moore, C. C. & Batchelder, W. H. (1996) in *Visualization of Categorical Data*, eds. Blasius, J. & Greenacre, M. J. (Academic, New York), in press.
27. Weller, S. C. & Romney, A. K. (1988) *Systematic Data Collection* (Sage, Newbury Park, CA).
28. Weller, S. C. & Romney, A. K. (1990) *Metric Scaling: Correspondence Analysis* (Sage, Newbury Park, CA).
29. SAS Institute Inc. (1988) *SAS/STAT User's Guide* (SAS Institute, Cary, NC), Release 6.03.
30. Nishisato, S. (1980) *Analysis of Categorical Data: Dual Scaling and Its Applications* (Univ. of Toronto Press, Toronto).
31. Greenacre, M. J. (1984) *Theory and Applications of Correspondence Analysis* (Academic, New York).
32. Gifi, A. (1990) *Nonlinear Multivariate Analysis* (Wiley, Chichester, U.K.).
33. Romney, A. K., Weller, S. C. & Batchelder, W. H. (1986) *Am. Anthropol.* **99,** 313–338.
34. Batchelder, W. H. & Romney, A. K. (1988) *Psychometrika* **53,** 71–92.
35. Batchelder, W. H. & Romney, A. K. (1989) in *Mathematical Psychology in Progress*, ed. Roskam, E. E. (Springer, Heidelberg), pp. 229–248.
36. Romney, A. K., Batchelder, W. H. & Weller, S. C. (1987) *Am. Behav. Sci.* **31,** 163–177.
37. Maher, K. M. (1988) Doctoral dissertation (Univ. of California, Irvine).
38. Weller, S. C. (1987) *Am. Behav. Sci.* **31,** 178–193.
39. Garro, L. (1986) *Am. Anthropol.* **88,** 351–370.
40. Iannucci, A. L. & Romney, A. K. (1994) *J. Quant. Anthropol.* **4,** 279–295.
41. Jameson, K. & Romney, A. K. (1990) *J. Quant. Anthropol.* **2,** 289–303.
42. Weller, S. C. (1983) *Hum. Organ.* **42,** 249–257.
43. Weller, S. C. (1984) *Am. Anthropol.* **86,** 341–351.
44. Weller, S. C. (1984) *Am. Anthropol.* **86,** 966–975.
45. Weller, S. C., Romney, A. K. & Orr, D. P. (1986) *Hum. Organ.* **46,** 39–47.
46. Klauer, C. & Batchelder, W. H. (1996) *Psychometrika*, in press.
47. D'Andrade, R. G. (1987) *Am. Behav. Sci.* **31,** 194–202.
48. Spearman, C. (1910) *Br. J. Psychol.* **3,** 271–295.
49. Brown, W. (1910) *Br. J. Psychol.* **3,** 296–322.
50. Nunnally, J. C. (1978) *Psychometric Theory* (McGraw-Hill, New York).