JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 8, 176–184 (1969)

A Psychological Study of the Semantics of Animal Terms[1,2]

NANCY M. HENLEY[3,4]

The Johns Hopkins University, Baltimore, Maryland

The semantic field of animal terms in English was investigated with five experimental techniques: free listing; pair ratings; triad ratings; verbal associations; and paired-associates learning. A set of 30 animals was scaled in the pair-rating experiment, a subset of six in one triad-rating experiment, and a subset of 12 in another triad-rating experiment. The same set of 30 animal names were stimuli for the associations. The subset of 12 were stimuli or responses in the paired-associate experiment. Results showed clear structure, with relevant dimensions of size and ferocity, for the animal terms. Furthermore, there was little effect of size of the set scaled, and different scaling techniques showed high correspondence. Different investigative methods gave highly similar clusters of animals. However, the paired-associate experiment was unsuccessful in eliciting a semantic space for animal names.

There have been many empirical studies of various semantic fields and relations, some psychological, some anthropological, and some linguistic (see, for example, Clark, 1968; Cliff, 1959; Deese, 1962, 1965; Howe, 1962; Jones and Thurstone, 1955; Mosier, 1941; Osgood, Suci, and Tannenbaum, 1957; Romney and D'Andrade, 1964). These have used mainly scaling and associative methods. Of these only those by Romney and D'Andrade and by Clark employed more than one technique. Therefore, there is little in the way of comparison among different techniques and their semantic implications.

The purposes of the research reported here are to explore the psychological techniques for the study of semantic fields in general and to characterize a particular semantic field. A further purpose of this research is to make comparison of the various psychological investigative techniques.

The semantic field investigated in this study was that of animal words in English, a field that both intuition and associative data (Deese, 1965) suggest to be highly structured.

The present study reports five experiments, each utilizing a different investigative technique. These were, in the order in which they were carried out, (1) listing of animals; (2) ratings of similarity and differences among triads of animals; (3) ratings of dissimilarities among pairs of animals; (4) associations to animal words; and (5) paired-associate learning, with animal names either as stimuli or as responses.

Two sets of animal words were used in the scaling, associations, and paired-associate experiments; one set consisted of the names of 12 common animals, and the other of a larger set of 30 animals. The set of 12 names, used in

[1] This research was reported in part at the Eastern Psychological Association meeting in Washington, D.C., 1968, and is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the Johns Hopkins University. It was supported by Public Health Service Grant MH-06550-06 and by Research Fellowship 1-FL-MH-35,977-01 from the Public Health Service.

[2] To obtain Tables A, B, C, and D referred to in the text of this article order NAPS Document 00247 from ASIS National Auxiliary Publications Service, c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001, remitting $1.00 for microfiche or $3.00 for photocopies.

[3] The author gratefully acknowledges the advice and encouragement of Dr. James Deese and Dr. Warren Torgerson. She also wishes to thank Richard Meltzer and Richard Degerman for their assistance with data analysis.

[4] Now at University of Maryland, Baltimore County, Baltimore, Maryland 21228.

the triad-rating and paired-associate tasks, was included in the larger set, which was used with pair ratings and associations. This design permits a comparison of size of set and of scaling techniques, though not independently.

All these experiments were carried out on the same set of 21 *S*s. In addition, another set of 21 *S*s performed a triad-rating experiment in which only six animal names were used, a subset of the 12. Nineteen additional *S*s made lists of animal names, and 27 additional *S*s gave free associations to animal names as stimuli.

Subjects

The 21 *S*s who were used for all experiments, and the additional ones who listed animals, were Johns Hopkins University summer session students, both male and female. They ranged in age from about 20 to about 45 years. All additional *S*s in other experiments were male Hopkins undergraduates recruited from the introductory psychology course.

EXPERIMENT I: LISTING

Method

Subjects were asked to list all the animals they could within 10 min.

Results

List length ranged from 21 to 110 animals, with a median of 55. Altogether 423 different animals were named. Of these 175 were named only once. Dog, lion, cat, horse, and tiger (in order of frequency) were all listed by more than 90% of the *S*s. Complete listing data are available as Table A from the National Auxiliary Publications Service.

Of the 29 animals listed by half or more of the *S*s, 26 (90%) are mammals. The 17 most frequently named animals are all mammals. However, as we pass to the less frequently named animals there are fewer mammals listed; only 27% of those named only once are mammals. As might be expected, the category *animals* is prepotently that of mammals.

Cohen, Bousfield, and Whitmarsh (1957) had *S*s list four examples for each of 43 different categories, of which one was that of four-footed animals. Though this category is more restricted than that given to the *S*s in the present experiment, the two sets of data are comparable, especially since those animals named most frequently in the present listing task are four-footed ones. A product-moment correlation was computed between log frequencies from the two lists. The resulting correlation is .65 (significantly different from zero, $p < .001$).

The frequency with which an animal is named is related to its frequency of use in the language in general. The Lorge magazine count (Thorndike and Lorge, 1944) was used as an estimate of the word-frequency count. For the set of 30 animal names, a product-moment correlation was computed with logarithmic transforms of the frequencies of the two lists, giving an r of .57 ($p < .01$).

There is reason to believe that in listing animals, people cluster names of animals which they consider to be similar. Bousfield (1953) has shown that they cluster semantically similar sets in free-recall experiments. If they do the same in free listing, then the differences between ordinal positions of similar animals would be less than the differences between ordinal positions of dissimilar animals. Thus an interanimal distance measure can be calculated from the listing data.

Based upon the data from the 21 *S*s who participated in all experiments, mean interanimal distances were calculated for the 12-animal subset in the following manner: (a) for each animal pair, for each *S*, the difference in order of listing the two animals was divided by list length and multiplied by 100; (b) mean distances were obtained by taking the mean over all *S*s for each of the 66 animal pairs.

The matrix of distances is shown in Table 1. The lowest mean distance is that between goat and sheep, 1.8; and the highest mean distance is that between cat and deer, 56.1. Other animal

TABLE 1
MEAN INTERANIMAL DISTANCES FROM LISTS

	Cat	Cow	Deer	Dog	Goat	Horse	Lion	Mouse	Pig	Rabbit	Sheep
Bear	47.2	27.7	40.1	49.6	19.1	29.0	22.6	29.5	21.4	20.3	16.1
Cat		30.9	56.1	2.0	29.0	25.3	24.1	24.8	43.0	41.5	47.1
Cow			43.6	30.2	11.0	7.7	24.5	34.1	17.0	27.9	8.2
Deer				50.9	44.5	43.0	44.7	39.9	41.1	19.9	53.1
Dog					17.0	24.0	26.9	27.5	45.0	39.4	46.8
Goat						7.2	23.1	39.6	19.5	21.8	1.8
Horse							28.6	32.6	25.7	30.1	15.2
Lion								33.2	29.3	33.3	35.0
Mouse									34.9	22.6	51.9
Pig										25.9	19.6
Rabbit											32.5

pairs with low distances are cat and dog, 2.0; goat and horse, 7.2; cow and horse, 7.7; cow and sheep, 8.2; and cow and goat, 11.0. Animals that are related psychologically do seem to be named in close proximity on lists.

EXPERIMENT II: PAIR RATINGS

Method

The stimuli in this experiment were the 30 animals of the large set. The method employed was the method of category ratings. The *S*s were presented with all possible pairs of the animals (435), one pair at a time and were asked to make a judgment of the amount of dissimilarity between the two, on a scale from 0 (no difference) to 10. At a second session one week later, *S*s made the judgments again, this time with the pairs in opposite (within-pair) order from that of the first session. A within-*S* reliability measure was obtained by the correlation of first and second judgments.

Individual differences among *S*s were examined before their mean dissimilarity matrices were averaged for scaling, by a procedure similar to the Tucker and Messick (1963) points-of-view method, followed by the application of Johnson's (1967) clustering technique (diameter method). The *S*s were found to divide into two groups and one odd *S*. The results of pair ratings will be presented only for the larger group of 18 *S*s who clustered together in the individual-differences analysis. The results for the group of 2 *S*s were found to be not too different from those for the 18-*S* cluster.

Dissimilarity matrices from all 18 *S*s were averaged to give mean pair dissimilarities as input to the TORSCA multidimensional scaling program (Young and Torgerson, 1967). This program performs repeated factor analyses of the scalar products matrix derived from monotone transformations of the original dissimilarities, followed by a series of rearrangements of the positions of the points. It arrives at a monotone transform of the original dissimilarities which is best from the standpoint of (a) minimum stress (Kruskal, 1964) and (b) agreement between derived distances and "best" monotonic transforms of the original dissimilarities (Torgerson and Meuser, 1962).

Results

Reliability coefficients for the 21 *S*s ranged from .29 to .85, with a median of .71. For the mean dissimilarity matrix for all *S*s the reliability coefficient is .97, indicating both that order of presentation of stimuli does not affect reliability, and that the *S*s showing low reliability in their ratings do not cause low reliability in the over-all matrix.

The values obtained for the stress index of the scaling in four, three, and two dimensions are respectively 6.8%, 9.4%, and 15.4%. The values for the index of agreement are .999, .998, and .994. The fit in three dimensions is judged acceptable here. Plots of Dimension 1 vs. Dimension 2, and Dimension 1 vs. Dimension 3, are shown in Figs. 1 and 2, respectively.

Examination of Fig. 1 indicates that the first dimension, on which elephant, giraffe, and camel are high and mouse, rat, and chipmunk are low, is a dimension of size. The second dimension has cow, sheep, and deer at

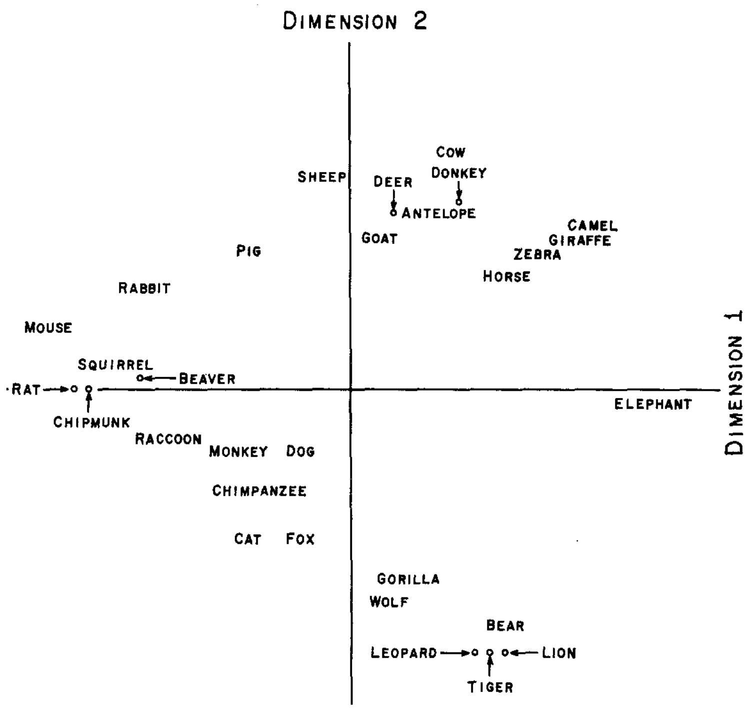

FIG. 1. Dimensions 1 and 2 of three-dimensional scaling of 30 animals.

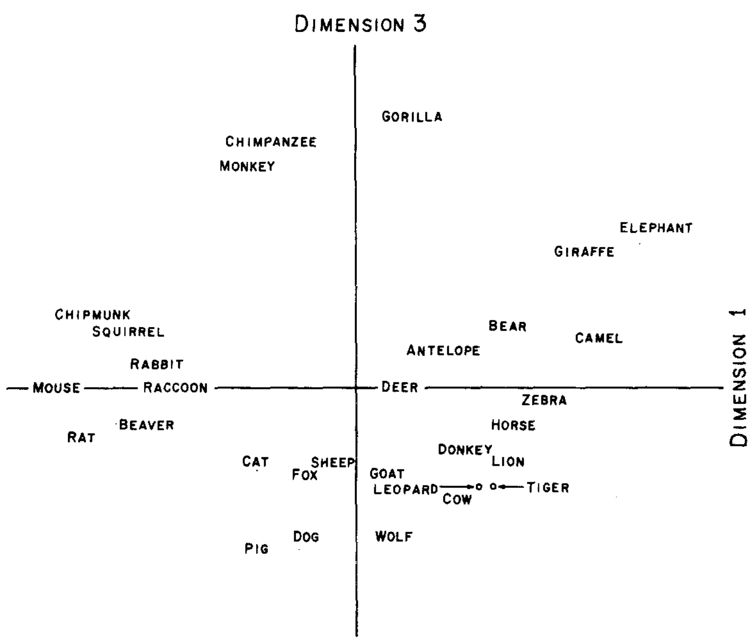

FIG. 2. Dimensions 1 and 3 of three-dimensional scaling of 30 animals.

one extreme and leopard, lion, and tiger at the other. It seems to be characterizable as one of mildness vs. ferocity. The third dimension, seen on the ordinate of Fig. 2, is harder to label. The three primates are high on this dimension, and it may, therefore, be one of resemblance or relatedness to man or something similar. (There is no reason to expect that a cognitive dimension need be labeled in the language.)

The animals group themselves roughly into four categories, one of small wild animals (mouse, rat, etc.) one of large, hooved herbivores (sheep, deer, etc.) one of primates (chimpanzee, monkey, gorilla), and one of other carnivores (tiger, bear, etc.). Pig is halfway between the small animals and the herbivores (which include many farm animals), and the elephant is in a class by itself.

The interanimal distance matrix from the listing task was correlated with the dissimilarity matrix from the pair ratings (for the 12-animal subset) with a low positive correlation, .17 (not significant), resulting.

Experiment III: Triad Ratings

Method

There were two triad-rating experiments, one with the six-animal subset and one with the 12-animal subset. For both experiments the method of triadic combinations (Torgerson, 1958) was used for the presentation of the stimuli to the *S*s. The *S*s were to judge which two animals were most alike and which two were most different.

All possible unordered triads were presented to each *S* for judgments, in two different (within-triad and between-triad) orders. Matrices of interstimulus similarities were obtained and converted to dissimilarities by subtraction from a constant.

For each *S*, a reliability quotient was computed. It represents the consistency with which he made his judgments on the two occasions on which he was required to judge the stimuli.

Results: Six-animal subset

An individual-differences analysis was made of the way the *S*s tended to rate the animals. Two groups were clustered, but since the dissimilarity matrices of these two groups correlated .78, it seemed reasonable to combine them for the final scaling analysis.

The reliability quotients for these *S*s range from .55 to .95, with a median at .85. As with the pair scaling, the TORSCA scaling program was used. Stress for the scaling in four, three, two, and one dimension, respectively, is 0%, 0%, 2%, and 12.6%. The index of agreement for four, three, two, and one dimension is 1.000, 1.000, 1.000, and .996. Two dimensions seem to provide an excellent representation. In this configuration, the first dimension seems to be one of degree of ferocity. Also there is grouping: lion and bear, both wild animals, are close on this dimension; cat and dog, both pets, are close; and pig and goat, farm animals, are also close. The attribute represented by the second dimension cannot be simply expressed.

The correlation between the pair- and triad-rating matrices for the same six animals is .90, $p < .001$, indicating both high agreement between scaling methods and little, if any, effect of context.

Results: Twelve-animal subset

Data from the same 18 *S*s used in the pair scaling were used in the triad scaling. Again, multidimensional scaling was done with the TORSCA program.

The reliability quotients for all 21 *S*s range from .59 to .88, with a median at .81.

The scaling program shows the values for Kruskal's index of stress for four, three, and two dimensions to be respectively, 3.0%, 4.8%, and 9.9%. Torgerson's index of agreement for four-, three-, and two-dimensional scaling is 1.000, .999, and .998. The data seem well scaled in three dimensions.

The first two dimensions that result from this triad scaling are the same as those resulting from the pair scaling. The first dimension, with horse and cow rated high and mouse and rabbit low, is one of size, and the second, with sheep at one extreme and lion and bear at the other, is one of degree of ferocity. Again there is the grouping into small animals (rabbit and mouse), herbivores (cow, deer, horse, goat, sheep) and other carnivores (bear, lion, dog, cat). The quality represented by the third dimension is ineffable.

The correlation of the matrix from the 12-animal triads experiment with that from the pair ratings is .93, $p < .001$; and that between matrices of the two triad experiments, .96, $p < .001$. Here the correlation is highest when the two methods are the same, and context is again seen to have little effect.

The correlation of the triad-rating data with those from the listing task (interanimal distances matrix) is .16 (not significant).

Experiment IV: Associations

Method

Subjects were instructed to respond with the first *animal* word which came to them when they read the stimulus word. The stimulus words were the 30 animal terms of the large set.

Results

A complete table of associations given is available as Table B from the National Auxiliary Publications Service.

Intersection coefficients (Deese, 1962, 1965) were computed for the 435 pairs of animals. These coefficients measure the elements of two sets (e.g., the associations to Animal 1 and the associations to Animal 2) in common. A complete table of intersection coefficients is available as Table C from the National Auxiliary Publications Service.

Of the 30 animals in the set used here, 17 are in common with the animals in the Cohen, Bousfield, and Whitmarsh (1961) associative data, for which the intersection coefficients are given by Deese (1965, Table 33, p. 149). The correlation between those data and the present is .63 ($p < .001$).

A clustering analysis was performed on the intersection coefficients by Johnson's (1967) hierarchical clustering technique. The following groups were obtained: chimpanzee, gorilla, monkey; beaver, raccoon, rabbit, chipmunk, squirrel; bear, elephant, giraffe, leopard, lion, tiger; mouse, rat, cat dog, fox, wolf; antelope, deer, cow, pig, goat, sheep, camel, donkey, horse, zebra.

The data were also subjected to a principal-components factor analysis, followed by a Varimax rotation. The resultant factor matrix for the first five factors is available as Table D from the National Auxiliary Publications Service. The first factor, on which beaver, chipmunk, rabbit, raccoon, and squirrel show high loadings, appears to be one of small wild animals, or "varmints." It is similar to a grouping obtained in the scaling analysis, and it corresponds to one of the groups of the clustering analysis of these same intersection coefficient data. The second factor has such animals as antelope, camel, cow, deer, donkey, giraffe, goat, horse, pig, and zebra showing high negative loadings. It is apparently a factor defined by medium-to-large size herbivorous animals. It corresponds to similar groupings in both the scaling of the rating data and the clustering of the intersection coefficients. The other factors can be characterized as (1) primates, (2) large carnivores (especially wild felines), and (3) small carnivores, respectively. These will be seen to be in exact correspondence with the clustering of these data, and in rough correspondence with the groups obtained in the scaling experiments.

For comparison of the intersection-coefficient data with data from the other experiments (listing and scaling), an intersection-coefficient matrix was made based on only the data of the same 21 *S*s who were used in those other experiments. The correlation of the intersection-coefficient matrix (similarities) with the matrix from the pair scaling (dissimilarities) is $-.70$ ($p < .001$). The correlation of the 12-animal subset from the intersection coefficient matrix with the 12-animal dissimilarity matrix from the triad-rating experiment is $-.59$ ($p < .001$). The correlation of this subset with the 12-animal matrix of interanimal distances from the listing task is $-.23$ (not significant).

Experiment V: Paired-Associate Learning

Shepard (1958) and others (e.g., De Soto and Bosley, 1962; Pollio and Draper, 1966) have presented evidence to show that stimulus and response generalization are related to distance in psychological space. More specifically, the confusions that *S*s make in learning paired associates are shown to be related to the psychological distances between members of the stimulus or response set.

In the present study, the subset of 12 animals was studied in paired-associate learning. The animals were used either as stimuli or as responses so that the confusions within the set could be investigated for structure they exhibit.

Method

For 11 of the *S*s, animals were the stimuli and the numbers 0 through 11 were the responses; for ten of the *S*s, the numbers were the stimuli and the animals were the responses.

Stimuli were presented on a memory drum at intervals of 4 sec, with the correct responses occurring 2 sec after the stimulus.

Results

The data for the condition in which animals were stimuli and for that in which they were responses were summed to give total number of mutual confusions between each animal pair. The matrix of animal-pair confusions shows little structure of the type expected. Its correlation with the 12-animal matrix from the listing experiment is .01; with that from the pairs experiment, −.08; with the triads matrix, −.07; and with that from the intersection coefficients, .04. Either the structure is entirely different from that elicited in the other experiments, or this experiment has not disclosed a psychological structure related to the cognitive structure of animal names.

The data were recast in the form of a matrix in which numbers were the row and column heads, and entries were viewed as confusions among numbers rather than confusions among animals. There are generally high entries on the line above the diagonal of this matrix, especially in the middle ranges. This fact would indicate the adjacent numbers were confused more often than distant numbers. Number pairs were reclassified into pairs one step distant (0 and 1, 1 and 2, etc.), two steps distant (0 and 2, 1 and 3, etc.), and so on up to 11 steps distant, and the mean confusions computed at each number of steps distant. The correlation of mean errors with steps distant is −.92 ($p < .001$).

When errors made with each number, either as stimulus or response, are plotted, the curve is typical of the serial position effect found in experiments where *S*s must learn serial lists. The data may thus be well accounted for by the assumption that *S*s were responding to the numbers as a serial list. The fact that the psychological distances between animals are not evidenced in the data is probably due to their being more, rather than less, discriminable than the numbers with which they were paired. Another strategy possibly leading to better results would have been to use as the associates terms not having any consistent structure, as did De Soto and Bosley (with men's names).

Discussion

Methodological Implications

All of the five experimental methods employed here have proved to be useful probes for exploring semantic structure, although the paired-associate experiment explored the structure of an unintended set (numbers). There is good correspondence between the associative and scaling methods. In addition, in spite of low correlations with associative and scaling matrices, the listing data seem to tend along the same lines. That is, there are low listing distances between certain pairs corresponding to low distances in the scaling data and high intersection coefficients from the associative data.

The finding that similar semantic structures are elicited by differing methods of investigation is important to research in semantics. There has often been a tacit assumption that there is a single structure for a given semantic field, irrespective of method, but there has been little empirical evidence supporting it. The results of the present experiment suggest that there is enough in common among different methods to support the view that a common semantic structure exists, irrespective of context and method, at least for such verbal items as were used here.

We may conclude from the intercorrelations among rating experiments that the method of category ratings (used with animal pairs) and the method of triadic combinations are interchangeable from the standpoint of results.

The high correlations, in the rating experiments, between dissimilarities obtained from different rating contexts (i.e., sets of 6, 12, and 30 stimuli) indicate that there is little effect of context. The differences among the correlations may be accounted for by the differences in methods and *S*s, in fact.

Semantic structure of animal terms

The supposition that the field of animal terms is highly structured is valid. The four experiments which successfully investigated the animal semantic structure found varying degrees of organization. The reliability measures used in the scaling experiments indicate that the semantic structure is in general quite stable (except for those few *S*s with low reliability indices).

The general picture of the semantic structure of the animal names for the *S*s of this study is one in which animals are preponderantly mammals, and in which mammals fit into a few fairly well-described categories. These categories are (1) small wild animals; (2) herbivores, including farm animals; (3) primates; and (4) other carnivores (which may be further subdivided into large and small, or zoo and and non-zoo animals). Mouse and rat fall with the first class in pair scaling, and with the last class in associative clustering. The elephant belongs to a separate class. The important dimensions in classifying these animals are those of size and ferocity.

Obviously, the set of animals used must determine to a certain extent just which classes and dimensions emerge. However, an attempt was made to sample across the domain of mammals. In addition, since context was shown to have little effect on relationships among animals in scaling experiments, it may be assumed that in a larger context these animals would still have the same positions in psychological space.

REFERENCES

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

CLARK, H. H. On the use and meaning prepositions. *J. verb. Learn. verb. Behav.*, 1968, **7**, 421–431.

CLIFF, N. Adverbs as multipliers. *Psychol. Rev.*, 1959, **66**, 27–44.

COHEN, B. H., BOUSFIELD, W. A., AND WHITMARSH, G. A. *The Connecticut free associational norms.* Technical Report No. 35. Storrs, Conn.: University of Connecticut, 1961.

COHEN, B. H., BOUSFIELD, W. A., AND WHITMARSH, G. A., *Cultural norms for verbal items in* 43 *categories.* Technical Report No. 22. Storrs, Conn.: University of Connecticut, 1957.

DEESE, J. On the structure of associative meaning. *Psychol. Rev.*, 1962, **69**, 131–175.

DEESE, J. *The structure of associations in language and thought.* Baltimore: Johns Hopkins Press, 1965.

DE SOTO, C. B., AND BOSLEY, J. J. The cognitive structure of a social structure. *J. abnorm. soc. Psychol.*, 1962, **64**, 303–307.

HOWE, E. S. Probabilistic adverbial qualifications of adjectives. *J. verb. Learn. verb. Behav.*, 1962, **1**, 225–242.

JOHNSON, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, **32**, 241–254.

JONES, L. V., AND THURSTONE, L. L. The psychophysics of semantics: an experimental investigation. *J. appl. Psychol.*, 1955, **39**, 31–36.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, **29**, 1–27.

MOSIER, C. I. A psychometric study of meaning. *J. soc. Psychol.* 1941, **13**, 123–140.

OSGOOD, C. E., SUCI, G. J., AND TANNENBAUM, P. H. *The measurement of meaning.* Urbana, Ill.: University of Illinois Press, 1957.

Pollio, H. R., and Draper, D. O. The effect of a serial cognitive structure on paired-associates learning. *J. verb. Learn. verb. Behav.*, 1966, **5**, 301–308.

Romney, A. K., and D'Andrade, R. G. Cognitive aspects of English kin terms. In Romney, A. K. and D'Andrade, R. G., (Eds.), *Transcultural studies in cognition. Amer. Anthropologist* (Special Publication), 1964, 146–170.

Shepard, R. N. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, **55**, 509–523.

Thorndike, E. L., and Lorge, I. *The teacher's word book of* 30,000 *words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

Torgerson, W. S., and Meuser, G. Informal notes on Torgerson and Meuser's IBM 7090 Program for Multidimensional Scaling. Cambridge: MITRE Corp., 1962. Mimeographed report.

Tucker, L. R., and Messick, S. An individual differences model for multidimensional scaling. *Psychometrika*, 1963, **28**, 333–367.

Young, F. W., and Torgerson, W. S. TORSCA, A FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behav. Sci.*, 1967, **12**, 498.

(Received June 24, 1968)