Memory & Cognition
1980, Vol. 8 (4), 354-366

Recall of semantic domains

DAVID C. RUBIN
Duke University, Durham, North Carolina 27706

and

MARTHA J. OLSON
Lawrence University, Appleton, Wisconsin 54911

The order of recall of lists of words learned incidentally was analyzed by multidimensional scaling similarity matrices based on the number of times words were retrieved next to each other. For the semantic domains of mammals, birds, and kinship terms, retrieval from very long-term memory, both for groups and individuals, and recall of recently learned lists produced multidimensional solutions similar to published solutions based on judged relatedness and associative overlap. For the squares of the Monopoly board and the names of the members of the Lawrence University faculty, for which clear a priori category structures exist, the form of clustering in the order and timing of recall that is commonly found in recall of lists learned recently in the laboratory was also found in the retrieval of lists learned incidentally through multiple exposures over long periods of time in the real world.

One way to obtain data about the structure of semantic memory is to ask people to list all the instances of a semantic domain that they can. As this task involves minimal external constraints, memory organization and retrieval strategies will determine the probability of any item being recalled (Battig & Montague, 1969), the time course of recall (Bousfield & Sedgewick, 1944; Gruenewald & Lockhead, 1980), and, most important for this paper, the order of recall. Because subjects need learn nothing new in the laboratory, the structure of semantic memory that results from real-world exposure to semantic domains can be studied directly. Because recall is not constrained, full advantage can be made of the open-endedness of the free recall task.

The approach suggested is not new. In 1944, Bousfield and Sedgewick asked subjects to name all the items they could from specified semantic domains. As Bousfield (1953) later noted, "Inspection of the data at that time revealed a significant phenomenon which, while apparent, appeared too elusive for quantification" (p. 229). Semantically related items appeared next to each other in the recall protocols, often occurring in rapid bursts. Because of the methodological difficulties involved, however, Bousfield and others abandoned the study of recall directly from very long-term semantic memory in favor of the more indirect, but easier to quantify, approach of having subjects learn lists consisting of words drawn from several distinct semantic domains (Kausler, 1974).

With recent advances in multidimensional scaling and semantic memory, Bousfield and Sedgewick's (1944) original problem is no longer too elusive for quantification. Given a similarity matrix, that is, a measure of association between all the members of a set of items, multidimensional spaces can be constructed. By using a similarity matrix based on Bousfield's (1953) measure of repetitions (i.e., the number of times two items are recalled next to each other), Bousfield and Sedgewick's original data could provide a multidimensional solution. Words often recalled next to each other would be considered similar and would therefore appear near each other in a multidimensional scaling solution. Anglin (1970), Friendly (1977), and Kintsch (1970) have demonstrated the usefulness of similar approaches in the study of recently learned lists. Here, emphasis is placed on the recall of the domains themselves, rather than on the recall of recently learned lists drawn from these domains.

Forming the next-to similarity matrix from free recall data is quite simple. A square matrix of zeros is constructed. The squares are labeled by listing alphabetically all the words recalled in an experiment along both the horizontal and vertical axis. As the next-to measure is symmetrical by definition, only the upper right-hand or the lower left-hand off-diagonal elements need be considered. Each time two words appear next to each other in a recall protocol, the value of the cell in the similarity matrix corresponding to the intersection of these two words is increased by one. For example, after

We wish to thank Jeanne Marini, Cindy Marriott, and Daniel Rebson for their help in collecting the data; Allan Collins, Herbert Crovitz, Joseph Fitzgerald, Michael Friendly, Davis Howes, Gregory Lockhead, Elizabeth Loftus, and George Mandler for their comments on the manuscript; Lance Rips for providing the similarity matrices from his study; and Donald Olivier for making available his hierarchical clustering program. The research was supported by a grant from Lawrence University. Part of this paper was read at the annual convention of the American Psychological Association, Toronto, August 1978. Reprints are available from David C. Rubin, Psychology Department, Duke University, Durham, North Carolina 27706.

0090-502X/80/040354-13$01.55/0

scoring the three recalls "cat, dog," "mouse, cat, dog," and "dog, cat, mouse," the three off-diagonal cells of the cat-dog-mouse by cat-dog-mouse similarity matrix would have the values of cat, dog = 3; cat, mouse = 2; and dog, mouse = 0.

The next-to similarity matrix based on free recall has several methodological as well as theoretical advantages over methods currently used to study semantic memory. On the methodological side, the free recall task given to the subject is open-ended: That is, the subject and not the experimenter chooses the items of the semantic domain. The items that the subject lists can therefore be taken as a definition of the semantic domain available to him at that time, whereas the items an experimenter chooses cannot. Thus, the problem of determining a priori which items belong to a semantic domain, when belonging is a matter of degree (Rosch, 1975), can be finessed. Moreover, context effects are imposed by the subject's semantic memory and recall strategies, not by the experimenter's choice of words.

Avoiding experimenter-chosen samples is advantageous in itself. Truly random sampling is difficult to perform without a definition of the population; in fact, in studying semantic memory, random sampling is often not even attempted. The practice of having the experimenter just decide upon a sample is, however, dangerous. For example, it is easy to imagine how to consciously choose a set of animal terms that would produce similarity judgments that would yield strong size and ferocity dimensions: In fact, King, Gruenewald, and Lockhead (1978) have done just that.

A second methodological advantage is that, unlike the situation with the use of similarity ratings, the analysis of the data is not obvious to the subject. Therefore, it is difficult for the subject to second-guess the experimenter, and likewise it is difficult for the experimenter to unwittingly influence the subject. A final methodological advantage is that the task given to the subject is both simple and reasonable. People who are reluctant or unable to judge the similarity of a pig and a bear may be willing to list all the animals they know. Thus, the technique may be easier than others to use with subjects from different cultures, as well as with young children and certain clinical populations.

On the theoretical side, the method provides an analysis that is easily interpretable in terms of existing theories. Most theories conceptualize semantic memory as a collection of hierarchically or spatially organized concept nodes (Friendly, 1977; Hutchinson & Lockhead, 1977). The recall of subjects can be viewed as a trace of their paths through their semantic spaces, one item being output each time an appropriate new node is encountered.

Various models can specify the process of tracing these paths. For instance, in terms of spreading activation theory (Collins & Loftus, 1975), activation is seen as spreading from the node currently being output to the nearest node in the semantic domain that has not recently been encountered. This new node is output and becomes the center for spreading activation. In terms of Shiffrin's (1970) search model, retrieval would continue from one subdomain of semantic memory until the number of draws needed to find a new node in that subdomain becomes too large. A new subdomain would then be entered. Thus, there are well specified relations between postulated processes of retrieval and the actual observed data.

While the method outlined has advantages, there is no measure of organization of memory that is independent of a performance task. Thus, the similarity matrices obtained here depend on subjects' search strategies, on the assumption that ordering in recall as measured by item-to-item association is an important index of internal structure, and so forth. It is precisely because each method introduces biases that additional, converging methods are useful.

Recently, Friendly (1977) has advanced another measure of similarities based on free recall data. The method has also been used by Henley (1969). Instead of counting the number of times words are recalled next to each other, as is done here, Friendly takes as his measure of similarity the distance in items between words normalized by the number of words recalled. Because Friendly argues for the superiority of his interitem distance measure, a brief comparison of these two measures is warranted. No attempt is being made here to argue for the ultimate superiority of one measure over another, but rather to argue that both the measure used by Friendly and the one adapted here from Bousfield's (1953) work are reasonable ways to quantify the invariance apparent in the free recall of lists learned incidentally over long periods of time in the real world. The quantification and study of such invariance are the major contributions of this paper; the particular method used is not.

The main theoretical advantage of the next-to measure is that the formation of the next-to similarity matrix makes only one classical assumption: Items recalled next to each other are associated in memory. In making this assumption, the technique not only ties current research in semantic memory to the work on clustering and to earlier work on associations and on associative structure (e.g., Deese, 1965), but also provides a measure that is interpretable in terms of models that postulate an item-by-item process (e.g., Collins & Loftus, 1975; Shiffrin, 1970). When the next-to measure fits easily into these other frameworks, the interitem distance measure must be derived from it.

In actual free recall data for semantic domains, there is a local cluster structure. On the average, items recalled near each other are more similar than items recalled farther away. This cluster size, however, is quite local and is prominent over only a small number of items, usually three or fewer (Gruenewald & Lockhead, 1980).

Whether an item is recalled 30 or 35 items away from another item makes little practical difference, depending more on which intervening subdomains are sampled than on the end items themselves. The next-to measure is empirically less than optimal because it drops off too sharply, counting all items not immediately next to each other as the same. The interitem distance measure drops off too shallowly, counting the difference between two sets of items separated by 0 vs. 5 items as the same as the difference between two sets separated by 30 vs. 35 items. Determining the function with the optimal drop-off would require either a more sophisticated theory than is currently available or a massive empirical effort. Fortunately, determining that function is not crucial, as the structure in the recall data is robust enough to be easily apparent with either measure. It may be noted that Friendly's (1977) comparison of the two measures used artificial data that lacked the local structure found in actual recalls.

Leaving this methodological diversion and returning to the main issue, what is to be attempted here is a description, in as theoretically neutral terms as possible, of the structure of semantic memory. This description will differ from most descriptions based on recall in that the recall is directly from semantic memory and is not complicated by being filtered through episodic memory for a recently learned list. The description will differ from most other approaches to the study of the organization of semantic memory in that recall rather than a judgment task is used as a basis for measuring structure. That is, this study employs the task that cognitive psychologists have studied most intensely and hopefully know the most about.

In pursuing this goal, comparisons will be made between the results obtained here and results obtained in laboratory learning of lists of words. Experimental psychologists in general and cognitive psychologists in particular are being urged to move out of the laboratory and explain real-world behavior (e.g., Neisser, 1976). The question is how to do so while maintaining some of the laboratory rigor and the techniques that have been developed. This paper can be seen as such an attempt. The words to be recalled here are semantic domains learned for the most part incidentally over repeated exposures in the subject's distant past. Comparisons with laboratory studies should help broaden the data base of psychology while hopefully indicating that in using simplified laboratory conditions, psychology has not removed all the interesting aspects of real-world behavior.

The strategy of this attempt is as follows. Initially, in order to provide a check on the method, recall from two semantic domains with clear a priori category structure will be analyzed using both the next-to measure and more standard measures of clustering. Next, the next-to measure will be applied to recalls of the three semantic domains most often studied by psychologists: mammals, birds, and kinship terms. The obtained similarity matrices will be compared with similarity matrices from individual subjects and with similarity matrices from the work of other researchers, who used a variety of methods.

EXPERIMENT 1

In an attempt to demonstrate clustering in recalls from very long-term semantic memory, two incidentally learned domains with clear category structures were used: the 36 distinct squares of the Monopoly board and the names of the 120 faculty members of Lawrence University. In the game of Monopoly, the squares of the board logically belong in sets. It was hypothesized that the Monopoly board would be clustered into these sets: in particular, the same color properties, railroads, utilities, draw cards, taxes, and corners. It was hypothesized that the faculty would be clustered by department. As the offices of most of the faculty were not arranged by department, both semantic domains used allowed a distinction to be made between a purely spatial form of clustering and the more abstract, conceptual form of clustering proposed. After using traditional techniques to measure the degree of clustering, the temporal course of the recalls and next-to similarity matrices will be examined.

Method

Subjects. Twenty Lawrence University undergraduates recalled the squares of the Monopoly board, and 52 last-term seniors from Lawrence University (4 majors from each department in the sciences and 2 majors from each department in the humanities) recalled the faculty.

Procedure. Subjects were asked individually to name either the squares of the Monopoly board or the members of the Lawrence faculty. Their recalls were tape-recorded. The recall period ended when the subject volunteered that no more items could be remembered.

Results

Recalled items were scored correct if they unambiguously referred to a proper referent. Thus, for Monopoly, the response "railroads" was incorrect because it had four possible referents, but "parking" was scored correct because it could refer only to Free Parking. All recalls were scored by two judges. There were only six cases of disagreements, which were decided by a third judge. Throughout the research, if the same response occurred more than once in a recall, only its first occurrence was considered correct. As such repetitions were rare, this arbitrary decision is of little practical importance. When a repetition of the same item did occur, the item usually appeared in a different semantic context (Gruenewald & Lockhead, 1980). In calculating the repetition and the next-to measure of clustering, and for similarity matrices, incorrect responses occurring between correct reponses were counted as separating the correct responses.

Interresponse times (IRTs) were measured between adjacent correct responses for all 20 subjects in the

Monopoly condition and for 1 subject from each of the 19 departments in the faculty condition. The taped sessions were recorded at 7.5 in./sec on a reel-to-reel tape deck and transcribed. The timing was accomplished by playing the tape back at 1.875 in./sec, thereby increasing the interval to be timed by a factor of 4. The average of three readings for each interval was used in the analysis.

Subjects recalled an average of 14.1 (SD = 6.7) squares of the Monopoly board and 36.7 (SD = 14.6) faculty members.

Clustering was measured using Bower, Lesgold, and Tieman's (1969) modified ratio of repetition [i.e., the number of repetitions of items from the same category/ (the number of items recalled–the number of categories recalled)] and Bousfield and Bousfield's (1966) formula for repetition expected by chance. For the Monopoly data, the modified ratio of repetition was .760; a value of .234 would be expected by chance [$t(19) = 8.64$, $p < .001$]. For the faculty data, the average value of the modified ratio of repetition was .849; a value of .133 would be expected by chance [$t(51) = 33.70$, $p < .001$]. As the modified ratio of repetition has a range of 1-0, the values of .760 and .849 represent a high degree of clustering in the recalls from both domains.

There is good evidence that clustering in the Monopoly recall was not due to the clustering of the geographic location of the squares on the Monopoly board, but rather to the role of the squares in the actual playing of the game. First, a perfect recall of all the squares in the order they appear on the board would yield a modified ratio of repetition of .307, well below the average value of .760 that was actually obtained. Second, categories with nonadjacent squares, such as railroads, utilities, draw cards, and corners were clustered as frequently as those with adjacent squares.

In Bousfield and Sedgewick's (1944) original data, related items often appeared to occur in rapid bursts of responding (Bousfield, 1953). In a series of experiments, Pollio has since verified this observation (Pollio, 1964; Pollio, Kasschau, & DeNise, 1968; Pollio, Richards, & Lucas, 1969). A similar result has been obtained here. In calculating the IRTs, the time between two items of the same category was considered as a within-cluster time, even if the two items constituted only an incomplete cluster. As the logarithms of the within and between IRTs were distributed normally, logarithms were used in all calculations. The geometric mean IRTs for within and between clusters were .85 and 3.98 sec for Monopoly and .93 and 5.50 sec for the faculty. Using the average $\log_{10}$ within- and between-cluster time for each subject, these differences were significant for both the Monopoly [$t(19) = 8.00$, $p < .001$] and the faculty [$t(18) = 26.47$, $p < .001$] data.

Pollio et al. (1969) found that in recalls from lists of recently learned items there is an increase of IRTs with position within a cluster. While this effect has not been consistently replicated with recently learned lists (Patterson, Meltzer, & Mandler, 1971), similar results were obtained here for recalls from very long-term semantic memory. As shown in Figure 1, there was also a trend for clusters of longer length to have shorter initial IRTs. No inferential statistics are available for these results because the observations were not independent; varying numbers of observations came from different subjects. Most subjects were probably aware that the clusters to be recalled here were of a definite size, and thus they may have tried to search each cluster until it was exhausted. For domains without clusters of a definite size, different results might be obtained (Gruenewald & Lockhead, 1980).

The analysis so far is very similar to what would be standard for lists of recently learned items, and quite similar results have been obtained. Before discussing the implications of these findings, similarity matrices of the Monopoly and faculty data will be formed.

Similarity matrices formed using the next-to measure were submitted to hierarchical clustering instead of multidimensional scaling techniques because of the demonstrated clustering of the data. That is, while the same conclusions would be drawn from multidimensional scaling techniques, they follow more directly from the clustering solutions. Several hierarchical clustering algorithms, including one based on the maximum similarity among the subclusters to be clustered (i.e., the nearest neighbor or connectedness method) and one based on the minimum similarity among the subclusters to be clustered (i.e., the furthest neighbor or diameter

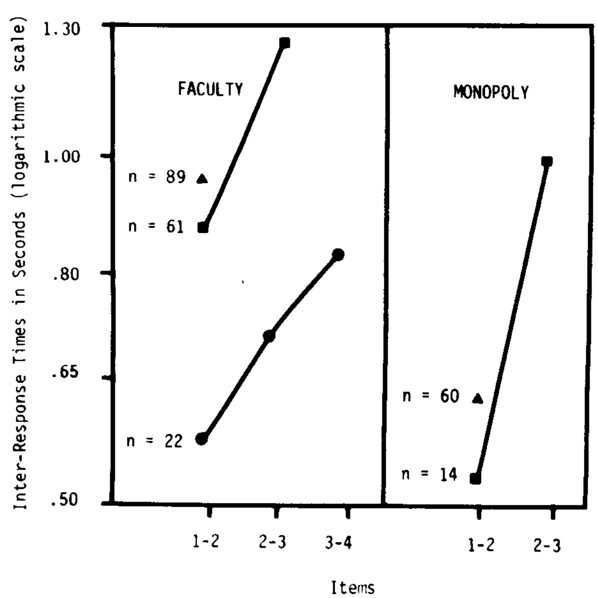

Figure 1. Interresponse times as a function of position within cluster for clusters two to four items long. Triangles indicate a cluster length of two; squares, three; and circles, four. The number of data points averaged for each cluster length is given by n.

method), were applied with similar results. Figures 2 and 3 show the hierarchical clustering solution for the algorithm based on the mean of the similarities between the items of the subclusters to be clustered. The horizontal position of the cluster in the figures is proportional to the strength of the cluster, with the strongest clusters to the left. In Figure 3, the faculty members have been labeled by their departments rather than their names. In order to meet computer limitations, the faculty data are displayed only for the 98 faculty members recalled seven or more times.

Figures 2 and 3 support the high degree of clustering found with the traditional clustering analysis. In the hierarchical analysis, however, no a priori categories are postulated, so that the method lends itself to the investigation of domains in which no single, clear, a priori category structure exists. The main structure in the similarity matrices was at the level of the categories postulated in the earlier traditional analysis. Structure at levels superordinate to this was quite weak, as indicated by the fact that such clusters are formed only at the extreme right of the figures.

Several exceptions to the structure postulated in the traditional analysis may be noted. In Figure 2, some associations not learned from playing Monopoly are apparent: Atlantic clusters with Pacific, Kentucky with Tennessee, and St. Charles with St. James, although they do not belong to the same category according to the a priori categories determined by the rules of the Monopoly game. In Figure 3, the lone member of the biology department clustered in the chemistry department has the same last name as the member of the

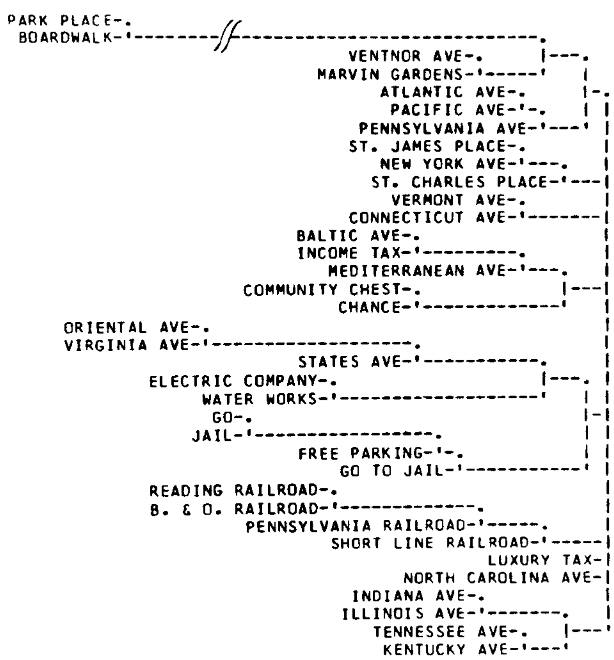

Figure 2. Hierarchical clustering solution for the Monopoly condition.

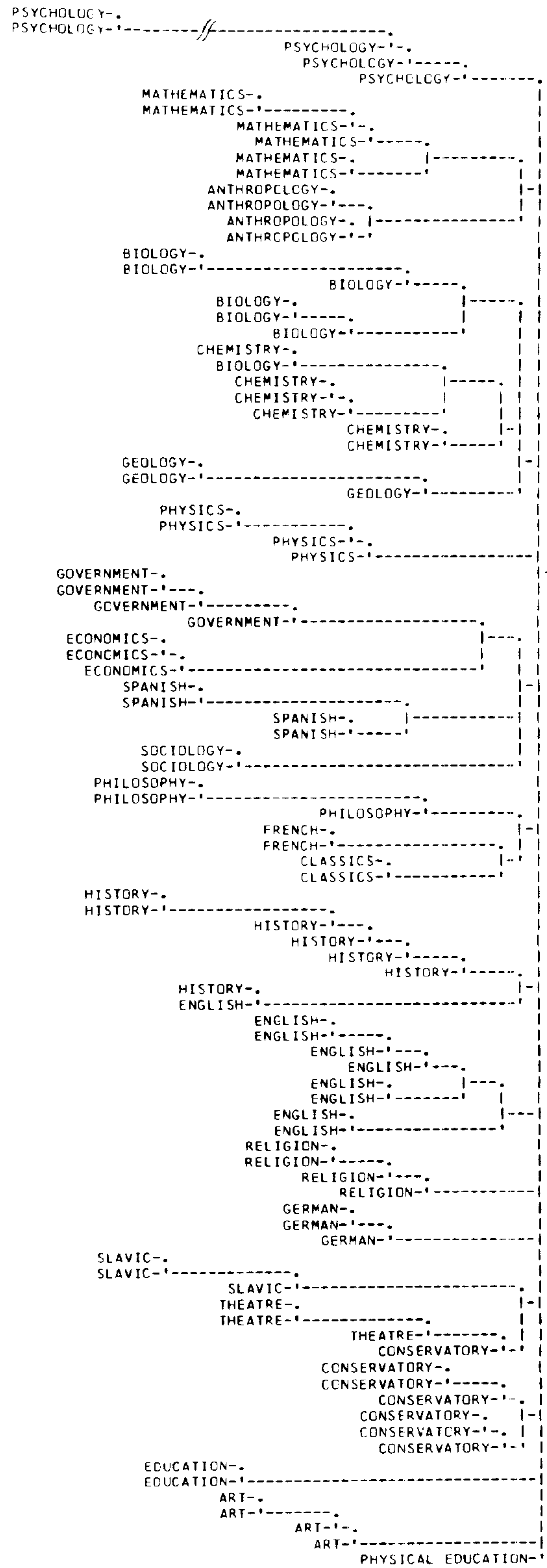

Figure 3. Hierarchical clustering solution for the faculty condition.

chemistry department he is next to. The same is true of the English professor clustered in History. Thus, some structure other than that initially postulated is evident using the more open-ended hierarchical clustering techniques. While to some extent this additional structure could be considered merely "clang" association, it could also be understood in terms of measures of semantic similarity, such as the number of common features.

Discussion

The traditional analysis of clustering demonstrates that clustering can occur in recall from very long-term semantic memory as well as from recently learned lists, thus extending results of laboratory list learning to lists learned incidentally over a period of time outside of laboratory controls. The only previous demonstration of clustering of items retrieved directly from very long-term semantic memory that could be found did not use a priori categories postulated for all subjects, but rather compared the categories into which individual subjects sorted words with the order in which those same individuals recalled the words (Dean, cited in Mandler, 1970).

For theories of semantic memory based on a spatial metaphor, the increase in IRT with position in the cluster indicates that items more central to a category or subdomain are more densely packed (or have shorter, quicker pathways). The decrease in initial within IRT with increasing cluster size indicates that categories that have a larger number of accessible items have their central items more closely packed than categories that have fewer accessible items. In fact, as priming would tend to cause items recalled later in a cluster to have shorter IRTs, the density differential must be great enough to overcome any effects of priming.

Shiffrin (1970) provides another simple interpretation of the timing data. Under this model, a subject would sample one word from the entire domain and then would sample words with replacement from within the category of that first word. If a novel word were found, it would be output. If a previously output word were found, another sampling would take place until a counter indicated that it was time to try another category. The increase in IRT with position in the cluster would be due to more items being sampled before a novel item was found. The decrease of the initial within-cluster IRT with increasing cluster size would be due to the lowered chance that the item sampled first would be sampled again before a novel item was found.

EXPERIMENT 2

Having demonstrated the feasibility of the next-to similarity matrix technique using two domains with clear category structure, we will examine three semantic domains for which subjects have no clear category structure. The three, which have all been previously studied by psychologists, are mammals, birds, and kinship terms.

Henley (1969) performed the most extensive study of the semantic domain of mammals, forming similarity matrices from interitem distance in free recall, pair ratings, triad ratings, and intersection coefficients from free associations. She found consistent structure that could be characterized either by claiming that the domain was determined by two dimensions, size and ferocity, or by claiming that the domain consisted of subdomains, or categories, labeled small-wild animals, herbivores including farm animals, primates, and other carnivores. The other carnivores category could be divided further, either into large and small animals or into zoo and non-zoo animals. Her findings are supported by Deese's (1965) study using intersection coefficients from free associations and by Rips, Shoben, and Smith's (1973) study using rating scales and reaction time measures.

The domain of birds has been studied by Rips et al. (1973). Using rating scales and reaction time measures, they found the same size and ferocity dimensions that Henley (1969) did for mammals. Their multidimensional scaling solutions also reveal a subdomain structure consisting of predators, wild songbirds, domestic pet-like birds, and domestic edible or farm birds.

Kinship terms have been studied extensively by psychologists and anthropologists (e.g., Fillenbaum & Rapoport, 1971; Miller & Johnson-Laird, 1976). Three basic dimensions underlie the domain: sex, a measure based on lineality, and a measure based on generation. Empirical studies by Fillenbaum and Rapoport (1971) support the three underlying dimensions (although they do not consistently appear as geometric dimensions on their multidimensional scaling solutions), as well as the following subdomains: immediate family, consisting of mother, father, sister, brother, daughter, and son; collateral family, consisting of aunt, uncle, cousin, nephew, and niece; and people two generations away, consisting of grandmother, grandfather, granddaughter, and grandson. Thus, all three domains chosen for study can be viewed as having both underlying dimensions and subdomain structures.

Method

Three groups, totaling 106 undergraduates, at the University of Wisconsin Center, Fox Valley were given 1 min/category to write down as many four-legged mammals, birds, and kinship terms as they could think of. The relatively short time interval was used, to follow other research in the area (e.g., Battig & Montague, 1969). The order of the domains was different for the three groups, with each domain appearing once first, once second, and once third. There was a brief rest period between the recalls of the different domains.

Results

Scoring our data according to Battig and Montague's (1969) rules, the correlations between the total number of times each item was recalled in Battig and Montague's and our data calculated over all the responses listed were r = .952 (n = 160) for mammals, r = .926 (n = 208) for birds, and r = .925 (n = 112) for kinship terms. Thus, in spite of the increase in recall period from 30 to 60 sec and regional differences in terms, the studies contain very similar data, and therefore it is likely that the analysis that follows would also hold for Battig and Montague's data.

Before forming the similarity matrices, the recall lists were condensed by scoring all responses that referred to the same object or person as the same. For example, mother, mom, mama, and ma all refer to the same person and thus were all scored as mother. Mammals and birds belonging to the same species were considered to be the same. Grzimek (1972-1975) and Walker (1968) served as references for mammals, and Thomson (1964) served as a reference for birds. There was one ambiguous case for the kinship terms: Cousin was assumed to the same person as first cousin.

In order to obtain a reasonable number of items for presentation, to keep calculations within reasonable bounds, and to remove items that were recalled so rarely that their next-to scores might be less reliable, only those items that were recalled by more than 10 subjects were analyzed. This resulted in 29, 30, and 24 items/domain. Similarity matrices based on the number of times these items were recalled next to each other were constructed. The resulting multidimensional scaling solutions using smallest space analysis (Lingoes, 1973) are shown in Figures 4, 5, and 6. Hierarchical clustering solutions, which are not shown, were in good agreement with these figures.

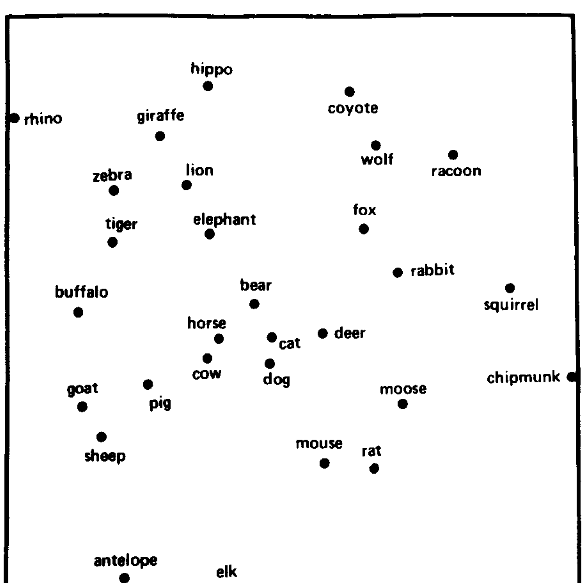

Figure 4. Smallest space analysis for the 29 mammals recalled by more than 10 subjects.

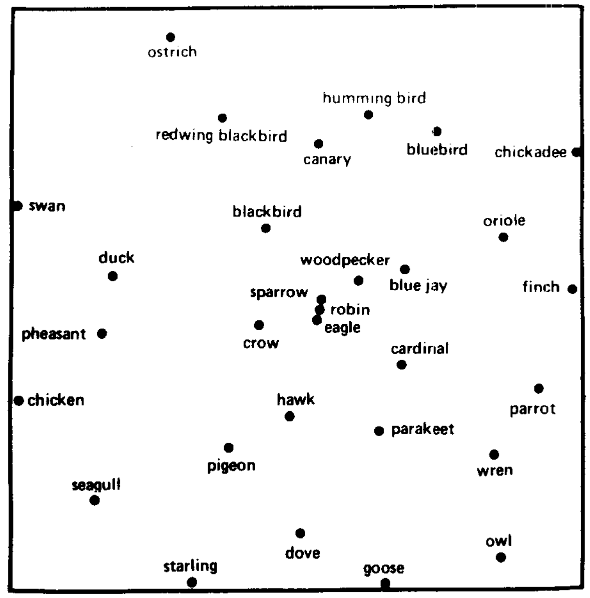

Figure 5. Smallest space analysis for the 30 birds recalled by more than 10 subjects.

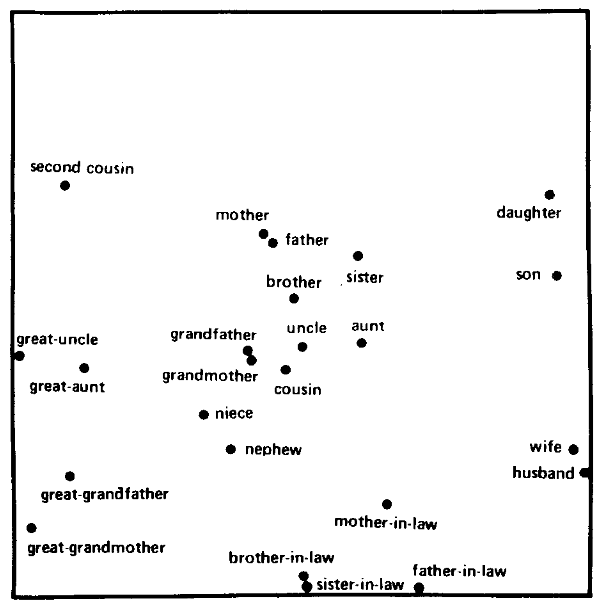

Figure 6. Smallest space analysis for the 24 kinship terms recalled by more than 10 subjects.

Smallest space analysis tries to match the rank order of the distances between the points on the figure to the inverse of the rank order of the similarity of items in the matrix. A measure of the proportion of failures is called the coefficient of alienation. It has a range of 0-1. For Figures 4, 5, and 6, the coefficients of alienation are .181, .245, and .158, respectively. In general, throughout the paper, increasing the number of dimensions from one to four results in a smoothly decreasing coefficient of alienation. For example, for one to four dimensions, the average values for this experiment were .339, .195, .140, and .101. As there were no clear breaks in the plots of the coefficient of alienation vs. dimensions and as examination of the three- and four-dimensional plots revealed no new insights, two-dimensional solutions were chosen throughout the paper, due to their ease of presentation.

Figure 4 reveals that, except for the displacement of deer and moose, the smallest space analysis fits Henley's (1969) subdomain scheme quite well. A horizontal size dimension with large mammals to the left and a vertical ferocity dimension with more ferocious animals to the top may also be seen, although there are many exceptions. The bird domain shown in Figure 5 is not as regular, but some of the same clusters that were noted in the Rips et al. (1973) solution are apparent. The size and ferocity dimension, however, did not appear.

The kinship terms shown in Figure 6 show the same subdomains as were found by Fillenbaum and Rapoport (1971) and an additional "by-marriage" subdomain consisting of wife, husband, and in-laws. The sex dimension is the clearest, although it fails to show up as a third geometric dimension; rather, kinship terms differing only in sex are located adjacent to each. A weak vertical lineality dimension and a weak horizontal generation dimension can be seen in Figure 6. If terms for grandchildren had been remembered more often by our subjects, they might have clustered with the grandparents terms, as Fillenbaum and Rapoport (1971) found, causing the generation dimension to become a generations-from-ego dimension.

Thus, there seems to be fairly good agreement among the findings of Experiment 2 and previous work with these three domains. A more detailed comparison will be presented later. In general, the subdomain structure provided a better description of the figures than did an underlying dimension structure. Here, as in Fillenbaum and Rapoport (1971), theoretically important dimensions for kinship terms displayed themselves in terms of subdomain clusters, rather than in actual geometrical dimensions.

Discussion

Retrieval from very long-term memory has been studied for five semantic domains using similarity matrices based on the number of times items were recalled next to each other. For the two domains from Experiment 1, which were learned relatively late in life and for which clear a priori subdomains exist, and for the three domains learned earlier in life, for which no such clear a priori subdomains exist, intelligible structure was found. While all semantic domains need not yield such simple results (Morton, 1975), the generality of the findings is encouraging. Three questions still need to be examined before much theoretical importance can be attached to these findings: (1) the relation of the retrieval from very long-term memory task used in Experiments 1 and 2 to the more commonly studied free recall of recently learned lists, (2) the relation of the grouped data presented in Experiments 1 and 2 to individual subjects' data, and (3) the relation of the similarity matrices obtained from the next-to measure used here to similarity matrices obtained by other methods.

EXPERIMENT 3

The studies presented so far all involve retrieval from very long-term semantic memory. Most work on organization in recall, however, is based on recently learned lists. In order to examine the relationship between these two tasks more closely, subjects were asked to learn a list of items drawn from each of the semantic domains used in Experiment 2.

Method

Subjects. Twenty-four Lawrence University undergraduates volunteered for the list learning task.

Procedure. Three lists were presented: one consisting of 12 mammals scaled by Deese (1965), Henley (1969), and Rips et al. (1973), one consisting of the 12 birds scaled by Rips et al. (1973), and one consisting of 11 kinship terms scaled by Fillenbaum and Rapoport (1971). A 12th term, self, was included to determine the subject's own place in the domain and to form a complete set of 12 terms. A 3-sec exposure was used for each word, yielding a total presentation time of 36 sec for each list. A blank slide was presented for a 12-sec unfilled retention interval, whose end was signaled by a slide asking the subject to "please write now." The delayed recall was used in an attempt to increase the degree of organization by allowing a period in which related items could be rehearsed together. The order of the presentation of lists was random, with the restriction that each list appear first, second, and third an equal number of times. The order of presentation of words within lists was random, with the restriction that each word follow each other word twice and occur in each serial position twice.

Results and Discussion

Subjects recalled an average of 10.1 (SD = 1.2), 9.9 (SD = 1.3), and 10.9 (SD = 1.0) items correctly from the mammal, bird, and kinship lists, respectively. Smallest space analyses performed on the next-to similarity matrices are shown in Figure 7. The coefficients of alienation for the three maps were .222, .215, and .226. In general, the maps are in agreement with those of the previous experiment. The underlying dimensions and subdomains were not as clear for the mammal domain as they were in the previous experiment. They were however, clearer for the kinship term domain, which revealed strong vertical generation and horizontal lineality dimensions. Some of these differences might have been introduced by the limited number of experimenter-selected items in each domain. For example, the items rabbit clustered with in Experiment 2 were not present in Experiment 3.

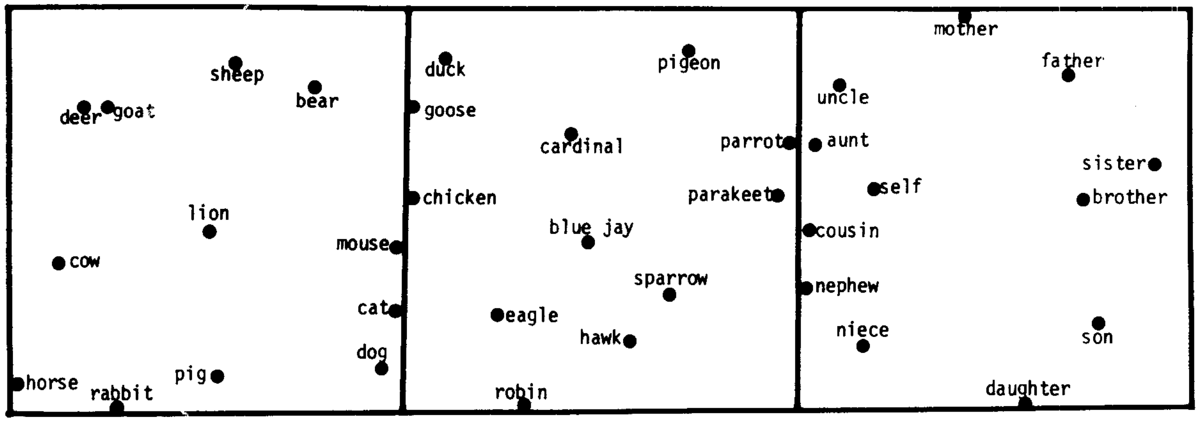

Figure 7. Smallest space analysis for the mammal, bird, and kinship term domains for the list learning experiment.

The subjects in the first two experiments were asked to retrieve familiar items from very long-term memory. In contrast, the subjects in this experiment had to remember which familiar items had appeared on a list. In Tulving's (1972) terms, the subjects in the first two experiments relied solely on semantic memory for retrieval, whereas the subjects in this experiment also had to rely on episodic memory. Experiment 1 demonstrated that the type of category clustering commonly noted in the timing and order of items recalled in episodic memory tasks was also present in a semantic memory task. This experiment, in conjunction with Experiment 2, demonstrates that the actual organization present in the order of the items recalled is similar. Thus, the first three experiments argue that the kind of results observed when both episodic and semantic memory are tapped can be obtained when only semantic memory is used. This is a comforting finding, considering the amount of research that has been based on the assumption that clustering in recently learned lists is evidence for organization in semantic memory. It also indicates that the distinction between semantic and episodic memory that has helped in conceptualizing this experimental literature may be of little help in accounting for it. Nonetheless, there are some differences between the particular multidimensional scaling solutions obtained in Experiments 2 and 3. More research would be necessary to decide if the differences are reliable and, if so, whether they are due solely to the inclusion of an episodic memory task at the time of retrieval.

EXPERIMENT 4

The results presented so far have all been based on group data. In this experiment, similarity matrices for individuals were obtained. In this way, not only could the results be shown not to depend on the grouping of the data, but also the magnitude and nature of individual differences could be assessed.

Method

Subjects. Nine Lawrence University undergraduates volunteered for the experiment.

Procedure. Every weekday evening for 4 weeks, the subjects were asked individually to write down as many four-legged mammals as they could in a 1-min period. Weekends were used to fill in missing trials, so that each subject provided 20 recalls. The subjects were instructed not to try to recall their previous recalls, but rather to recall the first names that came to mind.

Results and Discussion

An average of 16.3 (SD = 2.6) items were recalled per trial by the nine subjects. The 16 mammals that were recalled at least once by each subject were included in the analysis. The word "goat," which was recalled by five subjects, was also included so that the matrices of these five subjects would contain all of the mammal terms used in Experiment 3. Similarity matrices were formed. Any term that failed to be recalled next to at least 1 of the remaining 16 terms in an individual's similarity matrix was then removed in order to prevent the smallest space analysis from providing solutions with that term causing all other terms to be crowded into a corner. The results are presented in Figure 8. The coefficients of alienation ranged from .037 to .139, with a mean of .084.

In spite of the general agreement, individual differences are apparent in Figure 8. Not all subjects seem to have had underlying size and ferocity dimensions. Also, Subjects b, d, and g, and to a lesser extent, a, c, and e, had a distinction between African and American large wild animals, but Subjects f, h, and i did not. Subject i instead seems to have separated the large wild animals into those that are horse-like and those that are not.

One possible criticism of this experiment is that subjects might tend to produce stereotyped recalls, remembering what they said on the previous day rather than searching semantic memory. To check against this possibility, Pellegrino's (1971) unidirectional subjective organization measure was used to see if subjects were repeating words in exactly the same order on consecutive days. If stereotyped recalls were occurring from one day to the next, many exact repetitions of long sequences of words would be expected. If stereotyped recalls were not occurring, then only repetitions of short sequences of words should be common, the repetition of short sequences being due not to the recall of previous lists,

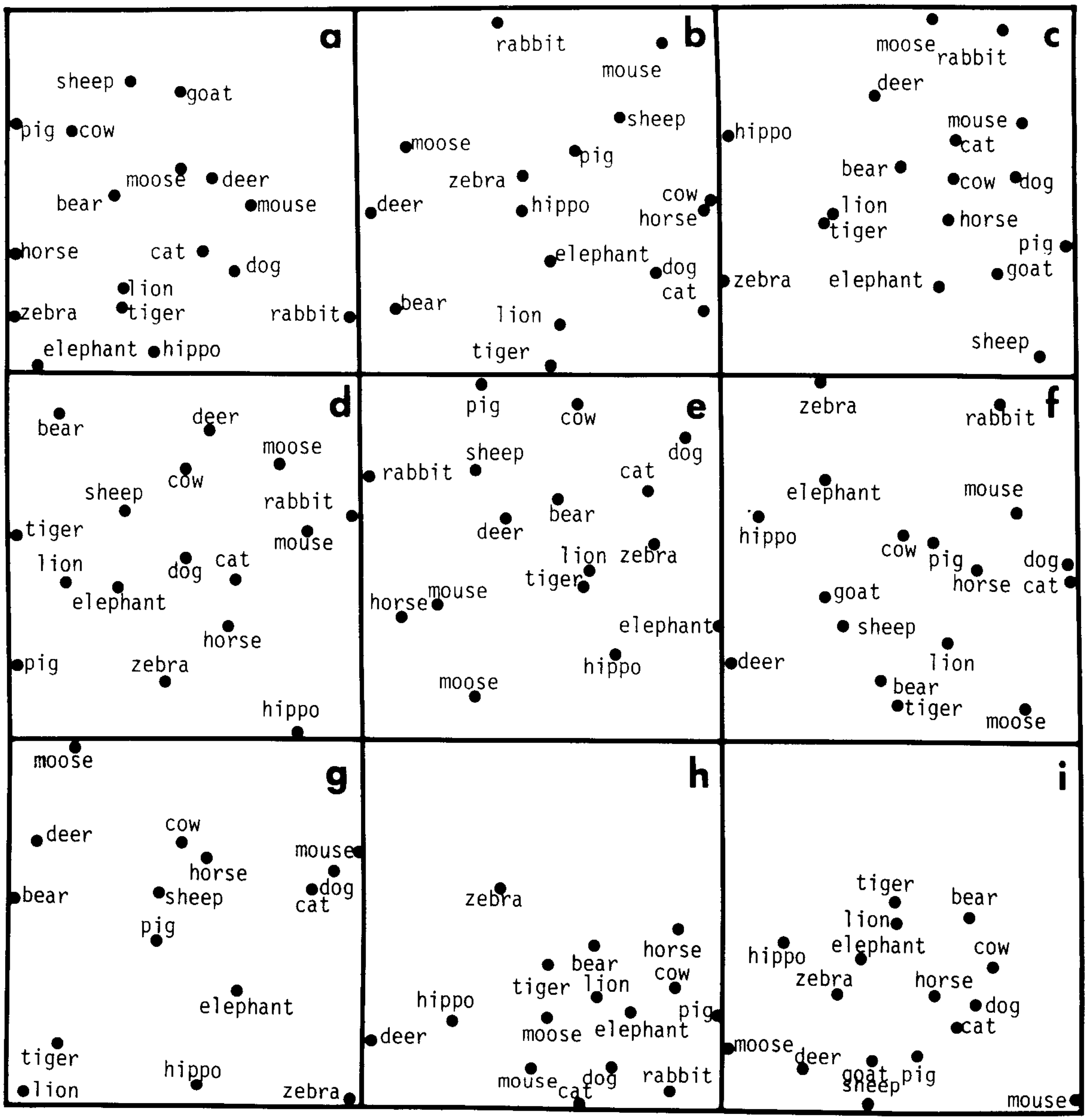

Figure 8. Smallest space analysis for nine individual subjects, a-i, for the mammal domain.

but rather to the stability of the associations in semantic memory on which the next-to measure is based. For sequences two, three, four, five, and six words long, there was an average of 2.59, .81, .25, .07, and .04 exact repetitions per recall. The number of observed repetitions of the longer sequences accounted for only a very small portion of the data, certainly not enough to indicate any form of stereotyping or to account for the structure in the similarity matrices.

COMPARISONS WITH OTHER STUDIES

The analyses in Experiments 1-4 were based totally on the order of items in recall. This section is an attempt to relate these findings to other measures of semantic memory. The studies presented in Figure 9 contained similarity matrices for the same 12 mammals. Figure 9 presents smallest space analyses for these mammals. The coefficients of alienation for Maps c-f are .158, .170, .130, and .180, respectively. All six maps are quite similar. With the exception of Map c, Henley's (1969) interitem distance measure, all show moderately strong size and ferocity dimensions, as well as farm animals (cow, goat, horse, pig, and sheep), large wild animals (bear, deer, and lion), and pet (dog and cat) domains. The two mammals that failed to enter into any clear subdomain (mouse and rabbit) tended to cluster with the pets, forming a loose subdomain of small animals. The fact that these maps do not exactly replicate those of Figure 4 of Experiment 2 is not surprising. Here, the 12 items of the maps were chosen by the experimenter. In Experiment 2, the 29 items were chosen by the

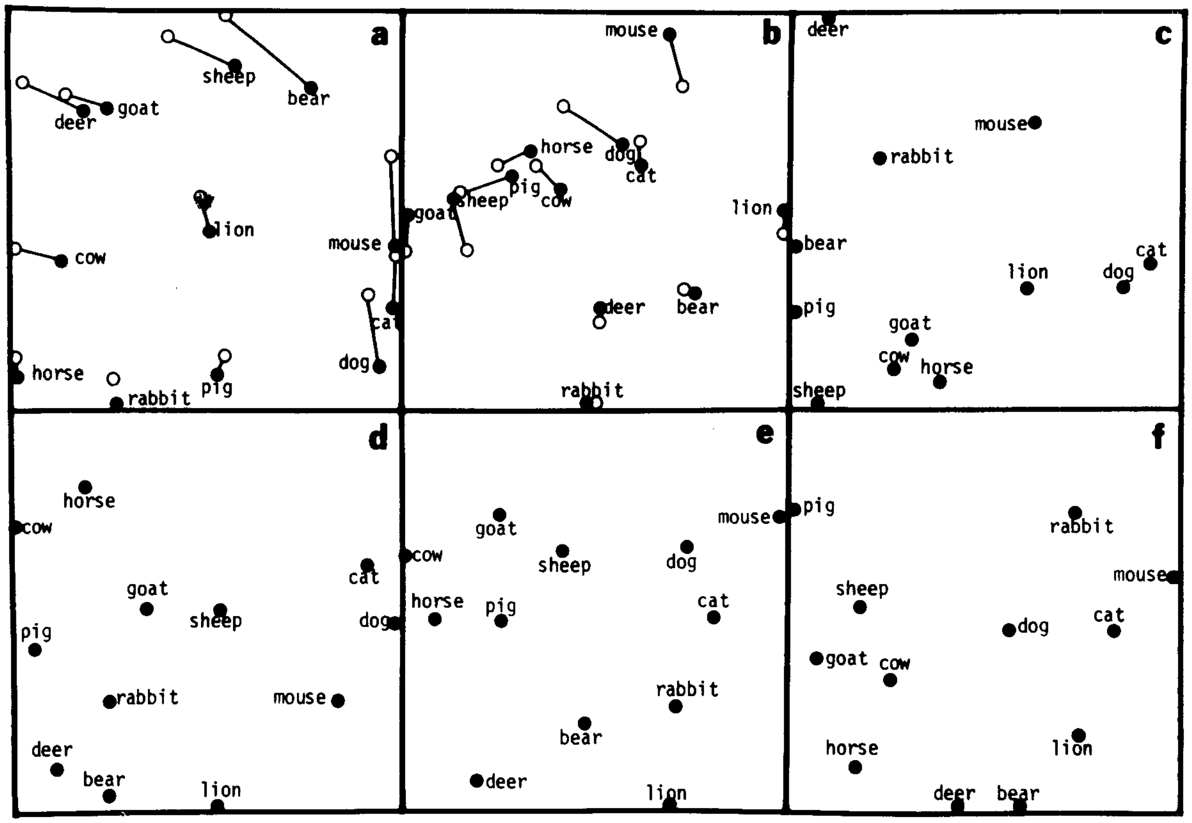

Figure 9. Smallest space analysis for the same 12 mammals based on similarity matrices from (a) the retrieval task of Experiment 2, (b) the list learning task of Experiment 3, (c) Henley's (1969) interitem distance measure, (d) Deese's (1965) intersection coefficient measure, (e) Henley's (1969) intersection coefficient measure, and (f) Rips, Shoben, and Smith's (1973) similarity ratings. For Maps a and b, the open circles are for the normalized matrices and the solid circles are for the nonnormalized matrices.

subjects. Several subdomains that might have appeared could not because some of the mammals belonging to these subdomains were eliminated. The relative positions of the 12 mammals in Figure 9 are, however, quite similar to their relative positions in Figure 4.

A more quantitative analysis is possible. Young and Lewyckyj (1979) provide a computer program, ALSCAL-4, that will fit many data matrices to the same multidimensional scaling solutions. The program indicates RSQ, the proportion of variance in each optimally scaled similarity matrix that is accounted for by the single multidimensional scaling solution. In order to make the analysis similar to those already presented, a two-dimensional nonmetric (ordinal) solution was used. The average RSQ for the six matrices of Figure 9 to a single solution is .452, compared with an average of .707 when six separate solutions are used. Thus, a single solution can provide a reasonable fit to all six matrices, although not as good a fit as six separate solutions. A single multidimensional solution to the 12 birds common to Experiments 2 and 3 and Rips et al. (1973) has an average RSQ of .691, compared with an average of .752 for three separate solutions. A single solution to the 11 kinship terms common to Experiments 2 and 3 has an average RSQ of .660, compared with an average of .780 for two separate solutions. Returning to the nine individual subjects of Experiment 4, a single solution has an average RSQ of .799, compared with an average of .973 for nine separate solutions. Given no evidence to the contrary, it can be assumed for now that all the methods tap the same aspect of semantic memory.

Figures 9a and 9b also demonstrate that the relative positions of items change little when the number of people who recall words next to each other is normalized by the number of people who recall those words in general. The open circles in these maps were obtained after applying Kintsch's (1970) method of normalization. Using ALSCAL-4, a single solution for both the normalized and nonnormalized matrices of Experiment 2 has an average RSQ of .708, compared with an average of .802 for two separate solutions. For Experiment 3, the figures are .685 and .683, respectively. Thus in terms of two-dimensional scaling solutions, normalization makes little difference. The decision not to include such a normalization in all figures was motivated partly by the empirical result that as long as words are recalled by approximately 10 or more people, there is little change in the conclusions drawn from the multidimensional scaling solutions. The major reason for not normalizing, however, was theoretical. The number of people who recall a word from very long-term, or semantic, memory is a good, if not a standard (Battig & Montague, 1969), measure of the degree to which a word belongs to a domain.

GENERAL DISCUSSION

The main advantage of free recall as a method of studying verbal behavior is that it allows organization imposed by the subject to be examined. Most studies of recall, however, structure the subject's free recall task in order to test for specific, predetermined forms of organization. Here, a more ethological attempt has been made to examine the organization in the free recall of semantic domains in as neutral terms as possible.

This approach led to several generalizations. First, clustering apparent in the order and timing of free recall data from recently learned lists can also be observed in retrieval from very long-term, or semantic, memory. Second, there is a structure to the sampled semantic domains that is similar for retrieval from very long-term semantic memory (for both individual and group data), recall of recently learned lists, associative overlap, and similarity judgments. Third, this structure can be described as a collection of subdomains and, in most cases, by underlying dimensions.

The underlying dimension claim, however, requires closer examination because it is the stronger of the two descriptions. That is, to claim that there is a subdomain structure is only to claim that certain terms are consistently found to be similar to each other. To claim that there are underlying dimensions, however, is to claim that the nature of that similarity is understood. The data presented here do not offer support for the stronger claim. First, the underlying dimensions do not appear consistently in the multidimensional scaling solutions; second, when they do appear, there are often many exceptions; third, in the case of the kinship term domain, while sex is clearly evident in subdomain clustering, it does not appear as a third dimension in three-dimensional scaling solutions. Thus, while the present findings, as well as those reanalyzed here, cannot argue that features do not exist for some semantic domains, in general they offer little support for that claim.

Practically, the results argue that the order of retrieval of items learned under controlled laboratory conditions is typical of the order of retrieval of items learned outside the laboratory. That is, laboratory controls were not necessary to observe the phenomenon of interest. To use Tulving's (1972) distinction, if order effects in retrieval are mainly due to semantic as opposed to episodic memory, then the exact nature of the episodic learning task inserted prior to recall, or even its complete absence, should have little effect on the order of retrieval from semantic domains.

The five semantic domains sampled here were each acquired incidentally, over a period of several years, without any experimenter control over, or even knowledge of, learning conditions or retention interval. They include domains that are among the earliest (e.g., mammals) and latest (e.g., facutly) semantic domains learned by the subjects. All, however, demonstrate retrieval effects similar to those found with lists learned recently in the laboratory. The protocols are even regular enough to allow for the examination of differences among individual subjects. Especially encouraging is the observation that models of processing constructed to account for laboratory observations, such as Shiffrin's (1970) search model or Collins and Loftus' (1975) spreading activation theory, can account for the kinds of data observed here.

REFERENCES

Anglin, J. M. *The growth of word meaning.* Cambridge, Mass: M.I.T. Press, 1970.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph,* 1969, **80**(3, Pt. 2).

Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports,* 1966, **19**, 935-942.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology,* 1953, **49**, 229-240.

Bousfield, W. A., & Sedgewick, C. H. W. An analysis of sequences of restricted associative responses. *Journal of General Psychology,* 1944, **30**, 149-165.

Bower, G. H., Lesgold, A. M., & Tieman, D. Grouping operations in free recall. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 481-493.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review,* 1975, **82**, 407-428.

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

Fillenbaum, S., & Rapoport, A. *Structures in the subjective lexicon.* New York: Academic Press, 1971.

Friendly, M. L. In search of the M-gram: The structure of organization in free recall. *Cognitive Psychology,* 1977, **9**, 188-249.

Gruenewald, P. J., & Lockhead, G. R. The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory,* 1980, **6**, 225-240.

Grzimek, B. (Ed.). *Grzimek's animal life encyclopedia* (Vols. 10-13). New York: Van Nostrand Reinhold, 1972-1975.

Henley, N. M. A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 176-184.

Hutchinson, J. W., & Lockhead, G. R. Similarity as distance: A structural principle for semantic memory. *Journal of Experimental Psychology: Human Learning and Memory,* 1977, **3**, 660-678.

Kausler, D. H. *Psychology of verbal learning and behavior.* New York: Academic Press, 1974.

King, M. C., Gruenewald, P., & Lockhead, G. R. Classifying related stimuli. *Journal of Experimental Psychology: Human Learning and Memory,* 1978, **4**, 417-427.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

Lingoes, J. C. *The Guttman-Lingoes nonmetric program series.* Ann Arbor, Mich: Mathesis Press, 1973.

MANDLER, G. Words, lists and categories: An experimental view of organized memory. In J. L. Cowan (Ed.), *Studies in thought and language.* Tucson: University of Arizona Press, 1970.

MILLER, G. A., & JOHNSON-LAIRD, P. N. *Language and perception.* Cambridge, Mass: Harvard University Press, 1976.

MORTON, J. Structuring experience—Some discussion points. In A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory.* New York: Wiley, 1975.

NEISSER, U. *Cognition and reality: Principles and implication of cognitive psychology.* San Francisco: Freeman, 1976.

PATTERSON, K. E., MELTZER, R. H., & MANDLER, G. Interresponse times in categorized free recall. *Journal of Verbal Learning and Verbal Behavior,* 1971, **10**, 417-426.

PELLEGRINO, J. W. A general measure of organization in free recall for variable unit size and internal sequential consistency. *Behavior Research Methods & Instrumentation,* 1971, **3**, 241-246.

POLLIO, H. R. Composition of associative clusters. *Journal of Experimental Psychology,* 1964, **67**, 199-208.

POLLIO, H. R., KASSCHAU, R. A., & DENISE, H. E. Associative structure and the temporal characteristics of free recall. *Journal of Experimental Psychology,* 1968, **76**, 190-197.

POLLIO, H. R., RICHARDS, S., & LUCAS, R. Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 529-536.

RIPS, L. J., SHOBEN, E. J., & SMITH, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior,* 1973, **12**, 1-20.

ROSCH, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General,* 1975, **104**, 192-233.

SHIFFRIN, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

THOMSON, A. L. (Ed.). *A new dictionary of birds.* New York: McGraw-Hill, 1964.

TULVING, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

WALKER, E. P. (Ed.). *Mammals of the world* (2 vols.) (2nd ed.). Baltimore, Md: Johns Hopkins Press, 1968.

YOUNG, F. W., & LEWYCKYJ, R. *ALSCAL-4 user's guide.* Carrboro, N.C: Data Analysis and Theory Associates, 1979.

(Received for publication September 5, 1979;
revision accepted January 9, 1980.)