Evidence of hierarchies in cognitive maps

STEPHEN C. HIRTLE
State University of New York at Albany, Albany, New York

and

JOHN JONIDES
University of Michigan, Ann Arbor, Michigan

Previous research suggested that the apparent hierarchical organization of landmarks in an environment will influence subjects' judgments about spatial characteristics of that environment. We extended this previous work to a natural environment that has no predetermined, well-defined hierarchical structure. Using an algorithm that generates a hierarchy of landmarks from recall protocols, we constructed hypothesized clusterings of landmarks for a set of subjects familiar with the space. Then we tested these hypothesized clusters in a series of tasks, all of which required judgments about distances in the space. The results of these tests suggest that subjects do cluster landmarks on the basis of nonspatial attributes, and that the clusters have consequences for performance in various tasks that require access to spatial information.

Create an image of the neighborhood in which you live. When you do this, some buildings and streets are included, whereas others are ignored. The selection of which landmarks to include in the image of a neighborhood is surely based on at least two criteria. One is the spatial proximity of the landmarks. Landmarks that are close together are likely to be represented together. But we speculate that your choice is based on more than just spatial proximity. We suspect that non-Euclidean information also influenced your selection of landmarks. A typical person's introspective image of Ann Arbor, Michigan, for instance, may have the farmer's market, the train station, and the Broadway bridge form a subjective group. To be sure, these three landmarks are not far from each other physically. But Euclidean proximity is not the only variable that may underlie subjective organization: The Bell Tower and State Street in Ann Arbor are as close to each other as the other three landmarks, but they are not subjectively part of the same neighborhood. These introspections lead us to hypothesize that cognitive maps represent not only spatial information, but also information about some nonspatial characteristics, such as subjective clusters of landmarks.

Of course, we need not rely solely on introspections to support this conclusion. There is evidence that implicates both the spatial and the nonspatial character of mental maps. On the spatial side, various experiments have shown that both distance and relative location are well-preserved in mental representations of physical space (Baird, 1979; Golledge & Zannaras, 1973; Jonides & Baum, 1978). Thus, some aspects of these representations may be adequately modeled by a two-dimensional space that is consistent with a Euclidean metric.

But further evidence indicates that these representations are more complex, since not all results can be predicted by a purely geometric model. For example, Cadwallader (1979) found three nonmetric results in the performance of subjects: (1) Rank orders of category judgments and magnitude judgments of distance correlated nearly .70 according to Kendall's Tau, (2) over half of the subjects were intransitive in at least 1 of 20 trials, when they judged which of two distances was larger, and finally (3) only 259 of 1,500 estimates were symmetric, with some asymmetries as great as 40%. Unfortunately, because Cadwallader did not provide baseline performance data for any of the tasks, it is not known whether the differences reported were consistent biases or random error. Still, the data suggest cause for concern about models whose representations of distance are based on purely spatial factors.

Further concern about such models is elicited when one explicitly recognizes the influence of cognitive variables in the representation of distance. Consider first the demonstrated effects of reference points. Sadalla, Burroughs, and Staplin (1980) found distance judgments between a reference point (a highly salient landmark) and a nonreference point to be asymmetrical, with distance from a reference point estimated as being smaller than the converse. Furthermore, Holyoak and Mah (1982) found that distances are judged differentially depending on their prox-

imity to a reference point. Specifically, distances near a referent are overestimated.

There are other cognitive effects on distance judgments. Sadalla, Staplin, and Burroughs (1979; see also Sadalla & Staplin, 1980a) found that a route containing high-frequency names is estimated as being longer than a route containing low-frequency names, but that prompting for the names eliminates the difference. Several studies have shown effects of familiarity, preference, and function on distance judgments (e.g., Briggs, 1973; Canter & Tagg, 1975), whereas others have shown that various incidental physical features of a space can alter judgments. For example, Sadalla and Magel (1980) found that routes with more turns are estimated to be longer than comparable routes with fewer turns. And Sadalla and Staplin (1980b) found that routes with more intersections were estimated to be longer than those with fewer intersections.

What we have, then, is an emerging picture of some spatial and nonspatial variables that play a role in the mental representation of physical space. This evidence rules out the possibility that an adequate model of a mental map can be built around a strictly Euclidean conception. As an alternative, consider a model proposed by Stevens and Coupe (1978), in which spatial areas are arranged hierarchically, so that judgments across clusters require knowledge of spatial arrangements within each cluster plus knowledge of the spatial arrangement of the superordinate structures. According to their model, processing in a spatial task is assumed to be a function of both spatial measurements and nonspatial (hierarchical) components. The evidence that favors this model comes from two sources: studies of hierarchical structure, and studies of the influence of barriers on spatial judgments.

Effects of Hierarchies

Direct evidence for hierarchical structure in cognitive maps comes from the experiments of Stevens and Coupe (1978). They observed that systematic errors occur in making directional judgments. For example, most subjects judge Reno, Nevada, to be northeast of San Diego, California, even though it is actually northwest. Similarly, Seattle is typically judged to be southwest of Montreal even though it is, in fact, northwest. In each case, according to Stevens and Coupe, the superordinate relationship produces incorrect judgments. For instance, in the first example above, subjects know that Nevada is east of California (perhaps by virtue of knowing that California has a coastline on the Pacific, whereas Nevada does not), and they use this knowledge about superordinate location to help in the judgments about subordinates. This bias, furthermore, occurs both for real-world locations and for those learned from artificial maps in the laboratory.

Further evidence of hierarchical clusters in cognitive maps can be seen in studies comparing within- versus across-cluster judgments. When measuring the time to verify directional statements, across-cluster judgments are made consistently faster than identically distant within-cluster judgments (Wilton, 1979). Furthermore, congruity

effects appear for within-cluster but not for across-cluster judgments (Maki, 1981). Finally, priming occurs for cities within states, but not for neighboring locales (Stevens, 1976). For example, a judgment about which of the pair Ohio and California is farther west facilitated a judgment about Cleveland and San Francisco, but not a judgment about Indianapolis and Reno. A spatial scanning model, unlike a hierarchical model, would have predicted that both pairs should be facilitated, given an appropriate control.

Extending the notion of clustering to a poorly differentiated area, Allen (1981) found that across-cluster distances were judged to be consistently longer than identical within-cluster distances as measured through a forced-choice task. However, the error rate was extremely high, with over 60% misclassification of distances at a 1:3 ratio. Thus, it is unclear if the results were due to an incorrect perception during the original learning or to an extremely strong clustering effect during the recall. In either case, the presence of boundaries altered the distance judgments.

To summarize, then, there are three basic clustering results. First, orientation judgments across clusters are both facilitated and biased by the superordinate relationship. Second, judgments can be facilitated by priming within a cluster. Third, distance judgments and the time to make distance judgments are influenced by cluster boundaries.

Effects of Barriers

A second source of evidence of clustering comes from studies on the effects of barriers. The primary dependent measure in barrier studies has been the judged distance between landmarks, either measured directly or derived from rank order data. Although the results of barrier studies are mixed, the general effect of barriers is that distances across barriers are overestimated in contrast to comparable distances that do not cross a barrier. Consider the following sources of evidence.

Kosslyn, Pick, and Fariello (1974) found that distance estimates between locations that crossed a barrier were exaggerated, when the distance estimates were extrapolated from rank order data. In a similar study, Newcombe and Liben (1982) also found exaggeration with rank order data but not with visual estimates. Cohen, Baldwin, and Sherman (1978) found overestimates across barriers within a summer camp when they used magnitude estimation, but found mixed effects when they used a reconstruction task. Thorndyke (1981) replicated the primary finding on a larger scale by demonstrating that "clutter" between cities increases estimated distances. All of these studies are consistent with a hierarchical model of the representation of space. All one need assume is that barriers create superordinate clusters whose distance from one another is exaggerated.

To summarize, then, one could argue that it would be a serious error to characterize the mental representation of physical space as a Euclidean survey map. Instead, there is experimental evidence showing that, in addition

to acquiring spatial knowledge of an environment, one also acquires information that is used to build a multilevel structure of the space. This structure includes hierarchies, reference points, distance knowledge, and semantic information about landmarks in the space (see also Hirtle, 1985).

The goal of the present experiment was to extend this model to account for the mental representation of spaces whose landmarks do not fall into well-defined clusters. Previous work by Stevens and Coupe (1978), Mak (1981), Witon (1979), and others has established the viability of a hierarchical model for spaces in which the hierarchies are unambiguous and are indicated clearly by obvious features in the stimuli (e.g., state boundaries). This was a useful beginning, since a failure to implicate hierarchical representations for such clear hierarchies would have cast serious doubt on the viability of this class of models. Having established this, though, it is appropriate to extend the application of these models to spaces in which a hierarchical organization may be suspected, but not clearly demonstrated.

In an important sense, our goal is much like the goal of memory researchers who have tried to implicate the use of organizational processes in the encoding, storage, and retrieval of verbal material. Here also, the first successful experiments included materials whose organizational structures were unambiguous (e.g., instances of items from a small number of well-defined categories). Later work then extended these early findings by developing methods of testing for subjective organization of randomly selected items, and by demonstrating that the structure of subjective organization had implications for performance in various tasks, such as free recall.

We have used this parallel between spatial and verbal memory as a heuristic to guide our present research. The first aim of our experiment was to measure subjective clusters for a real space whose cluster organization was not obvious. Then we collected data from performance tasks that required absolute and relative judgments of distances to test whether the subjective organizations that were uncovered had psychological reality as representations that were used in various tasks.

METHOD

Background About the Ordered Tree Algorithm

General description. Before describing the methodology of our study in detail, including the performance assessments of clustering, we find it worthwhile to review the procedure by which we assessed subjective organization (McKeithen, Reitman, Rueter, & Hirtle, 1981; Reitman & Rueter, 1980). Although this procedure has proved to be successful in revealing underlying organization for verbal material, its use with spatial memory is novel.

The technique was developed to explain regularities in free-recall data, in contrast to earlier clustering algorithms based on similarities (e.g., Johnson, 1967). In the context of free recall, subjects carefully memorize a list of items and then recall the items repeatedly, each time starting with any item and proceeding in any order. The input to the analysis algorithm consists of these recall protocols. The algorithm's output is a structure known as an ordered tree, an example of which is shown in Figure 1. The tree represents clusters
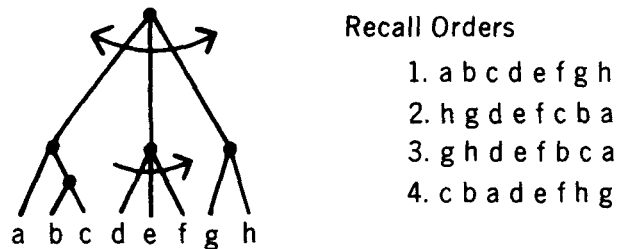

Recall Orders
1. a b c d e f g h
2. h g d e f c b a
3. g h d e f b c a
4. c b a d e f h g

Figure 1. Tree diagram for < (a(bc)) [def] (gh) >.

that the algorithm has uncovered from the recall orders. The clusters may be of one of three types: (1) unidirectional, in which items are recalled in only one order, (2) bidirectional, in which items can be recalled in one order or its inverse, or (3) nondirectional, in which there is no restriction on recall order.

The rationale for the analysis algorithm is to recover an internal structure that is modeled by an ordered tree. The basic assumption is that the obtained recall orders are a sample of all possible orders that may have been produced by a particular tree given that information can be properly represented by a tree and that all the items from one cluster are recalled before the subject moves on to another cluster.

The ordered tree algorithm proceeds by examining the recall strings for items that are recalled contiguously, regardless of order. These clusters, both discrete and overlapping, are then written into a lattice, and, finally, the lattice is written as a tree (the tree can also be represented as a series of parenthetical expressions, where square brackets {[ ]} indicate unidirectionality, angle brackets { < > } indicate bidirectionality, and parentheses {( )} indicate nondirectionality). The structure within a lattice can imply that a cluster is either nondirectional or directional. Of the directional clusters, to distinguish bidirectional from unidirectional, the individual strings are examined to determine whether both orders or only a single order actually occurs. In the example of Figure 1, DEF is unidirectional, with the order indicated by the arrow.

There is one final note to offer on the algorithm. A variety of strings are needed to adequately sample the tree. In order to induce variety, subjects are cued with various items. That is, they are asked to recall all items, but to start with the cue and those that go with it. Because cuing presumably disrupts the tree structure, the algorithm examines only the uncued portion of cued trials for both clusters and ordering information.

Application to cognitive maps. Previously, the ordered tree algorithm has been applied to memory organization for common words, computer programming keywords, physics symbols, and college course keywords (McKeithen, Reitman, Rueter, & Hirtle, 1981; Reitman & Rueter, 1980). In order to apply the algorithm to cognitive maps, we required data consisting of linear orders of landmarks in the space. We obtained these in much the same way that previous experimenters who used the algorithm had: Subjects were asked to memorize and recall landmarks in central Ann Arbor consisting of 24 buildings, 2 plazas, 4 streets, a bridge, and a river, for a total of 32 landmarks. These are shown in Figure 2. Subjects were then asked to recall the landmarks a total of 14 times, for which 10 of the recall protocols were cued with various landmarks.

Unlike the states or countries previously studied, the area studied has no predefined hierarchies or strict boundaries. However, several intuitive regions can be identified: a Central Campus area surrounding the Diag; a downtown area centered on Main St.; a Medical Center surrounding the University Hospital; and the Farmer's Market area.

Although most Ann Arbor residents would recognize these neighborhoods, the exact number of neighborhoods recognized and the corresponding boundaries fluctuate with individuals. Liberty Plaza is seen by some as being properly included in the downtown section and by others as being near Campus. CCRB, a recreation build-
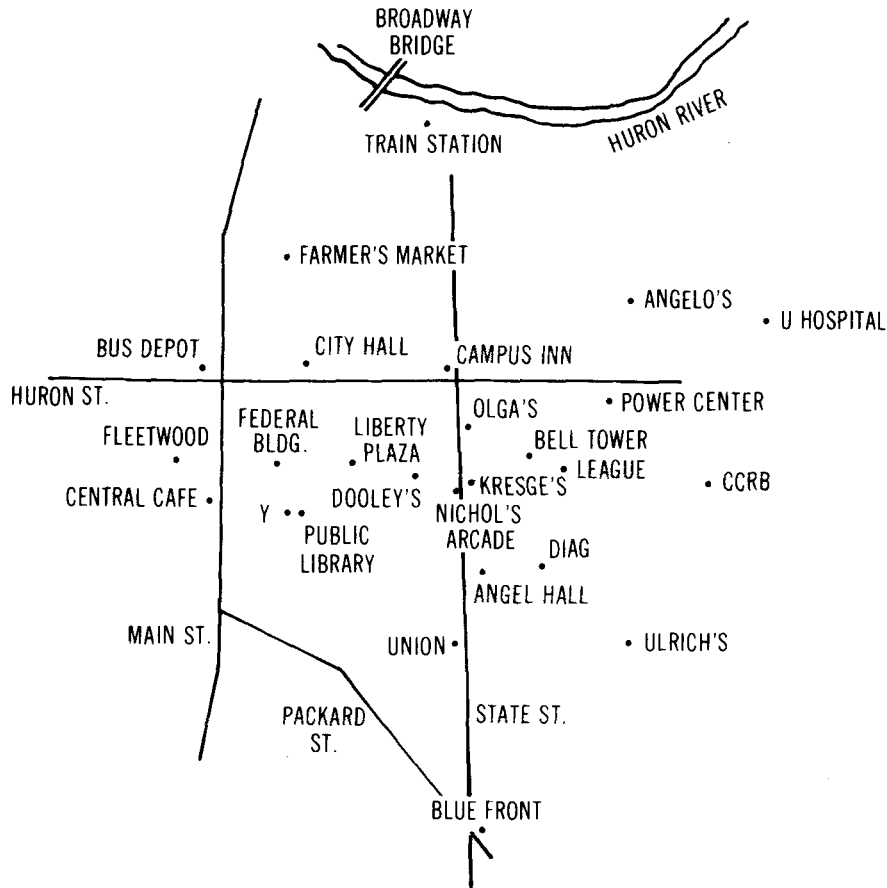
Figure 2. Map of central Ann Arbor indicating the relative locations of landmarks.

ing, is considered by some to be in the Medical Center and by others to be in the Central Campus.

In addition, where one lives or spends time influences the perception of space. Medical students might represent the University Hospital in a way very different from that of nonmedical students. Those who use a car as their primary mode of transportation segment areas differently from those who primarily walk. Thus, it is crucial that the analysis be performed individually, defining a separate tree for each subject.

Design

The subjects participated in a series of tasks. Day 1 of the experiment consisted of having subjects memorize the stimuli, produce the free-recall protocols, and construct maps of the landmarks. Day 2 consisted of a speeded classification task and a nonspeeded distance-judgment task. The subjects were run individually on both days. The subjects returned 6 weeks after the first pair of sessions to repeat the free-recall task in order to assess the stability of the clusters.

Subjects

Six female University of Michigan undergraduates (seniors) and graduate students, ranging in age from 21 to 30 years, participated in the experiment. All subjects had spent at least 2.5 years on the Ann Arbor campus and had used walking as their primary mode of travel in and around the campus and downtown areas. The subjects, each of whom was pretested for knowledge of buildings in the town, were chosen from a voluntary subject pool and were paid $14.00 for a total of 4 h of participation.

Stimuli

Memory task. The landmarks shown in Figure 1 were generated by asking a set of judges to name all prominent landmarks in

Ann Arbor. The final 32 were chosen to span four subjective neighborhoods: Downtown, Central Campus, Farmer's Market, and the Medical Center. These neighborhoods overlap to a large extent and were chosen only to provide a heterogeneous space. We neither tried to induce these neighborhoods, nor did we later uncover consistent clusters among these four groups. The subjects were pretested on a subset of these buildings, along with a few landmarks not part of the recall set, in order to assess their knowledge of Ann Arbor.

Distance judgment tasks. Triplets of points were created by assigning two anchors to a target, such that the distances to the anchors from the target were approximately equal and such that the anchors were in different directions from the target. Thus, two matched pairs were created from each triplet (target to Anchor 1, and target to Anchor 2), with the three points forming an isosceles triangle. The 64 pairs thus classified included distances that ranged from 173 to 895 m. The short pairs, of which there were 28, were less than 427 m, and the long pairs, 36 in number, were greater than 527 m. The standard that served as the criterion to discriminate short from long pairs was defined as the distance (466 m) between two particular buildings that were located in the center of the space, but that were not part of the recall set. Another 8 pairs, meeting the above distance constraints, were added to the 64 key pairs to form a test set of 72 pairs. An additional 48 pairs constituted a practice set. The practice and test pairs were randomized separately. The sets were randomized twice, once for the relative distance judgment task and once for the distance estimation task. The randomization occurred with the constraint that no landmark appear in the same position on consecutive trials.

Procedure

Pretest. The subjects were first asked to fill out a questionnaire that was used to determine their familiarity with a list of landmarks

in Ann Arbor. These data were collected for future studies and were not analyzed. Next, the subjects were shown individual color photographs, in which any identifying signs had been blackened, and were tested for knowledge of the landmarks used in the experiment. Each photograph had a red X indicating a building that the subjects had to identify. The X also marked a location on the building that was to be used as the referent for later distance judgments that subjects made. In each case, the X marked the main entrance, if there was one, or the center of the building, if there was not a main entrance. The subjects were required to identify correctly all buildings in order to continue with the task.

Memory task. The subjects were given 3×5 cards with the name of a landmark on each one. They were instructed to memorize the 32 stimuli with two goals in mind: to be able to recall each item and to be able to sketch a map locating each item. It was stressed that no omissions would be allowed during recall or map drawing. The subjects were instructed to spread the stimulus cards out on the table for memorization, and they were told that sorting the stimuli alphabetically would be disadvantageous for map drawing.

Free recall. The free-recall session consisted of 14 trials, 10 of which were cued with preselected cues. The subjects were first required to complete two consecutive perfect free recalls before beginning the cued trials. The remaining two free recalls occurred on Trials 8 and 14. On cued trials, the subjects were instructed to start recall with the cue and those landmarks that went with it, recalling all the objects in the recall set.[1] The cues were presented verbally. All recall protocols were recorded on cassette tape.

Map drawing. Having completed the free recall, the subjects were requested to draw a map from memory four times, starting wherever they wished on a blank sheet of 8½ × 11 in. (21.6 × 27.9 cm) white paper. Finally, the subjects were requested to construct a map of only the 26 buildings and plazas, on a 91-cm-square sheet of paper, using wooden blocks, on each of which was printed the name of a landmark. The construction task differed from the map-drawing task in both reducing the memory load and allowing easy alterations in placement.

Relative distance judgment. For this portion of the experiment, the subjects sat before a CRT controlled by a PDP-11/34 computer. A total of 120 pairs of landmarks, including the 64 key pairs, was tested. The subjects classified distances as being close or far as compared with a standard that was the distance from the Rackham Building to the Graduate Library (466 m). Pairs were presented individually in the following way: The word READY appeared for 1 sec, followed by a .5-sec pause, and then the name of the first landmark. After 4 sec, the name of the first landmark disappeared. There was a .5-sec pause, and then the name of the second landmark appeared. The subjects were instructed to respond quickly, but accurately, after the second landmark had appeared. The response CLOSE was given by depressing a key with the left index finger, and the response FAR was made with the right index finger on another key. The next trial began after a 4-sec intertrial interval. If an incorrect response was made, the first 1 sec of the intertrial interval contained the message ERROR. Immediate error feedback was given as an incentive to attend to the question. Cumulative feedback about accuracy and speed was given after every 12th trial. The pause between blocks of 12 trials also provided subjects with a self-paced rest period.

Distance estimates. The same 64 key pairs, plus 28 practice pairs, were presented individually on the CRT screen. Both names appeared on the screen, followed by an entry space for the answer. The subjects typed a response from 1 to 100 (indicating a distance magnitude estimate) on a numerical keypad, and then pressed a return key to enter the response. Trials were subdivided into 20 trial blocks, interrupted by rest periods. The subjects were under no time constraints for this task.

RESULTS

As discussed above, the experiment had two goals:

(1) to infer individual subjective organizations for landmark clusters using the ordered tree algorithm, and (2) to provide evidence from the performance tasks that could test whether the inferred subjective organizations play a role in a variety of actual judgmental situations. We shall report the results for each of these aims in turn.

The Ordered Tree Algorithm

Original testing. The free-recall data were used to generate trees individually for each subject. This raises a sampling problem, however. The total set of recall protocols that each subject generated are but a sample of the total population that could be generated from any tree, in many cases a small sample. Consider Figure 1, for example. There are 16 possible recall orders that can be generated from this tree, four of which are illustrated. It is possible, because of such factors as lapses of attention, that a subject may occasionally recall a string that cannot be generated by a particular tree, even though that tree may actually be a reasonable model of that subject's memory representation. Unfortunately, the ordered tree algorithm, being deterministic, has no patience for this sort of thing. It will always construct a tree that is consistent with all recall protocols of a particular set. We implemented a procedure, jackknifing, to guard against this intransigence of the algorithm. Under this procedure, each recall string was individually deleted from the set of strings of a subject. Then a new tree was constructed, and we calculated the total number of unique strings that it could possibly generate. This number (actually its logarithm, called PRO by Reitman & Rueter, 1980) was compared with the comparable figure when all strings were included in the original set of recall protocols. If there was a substantial difference either in PRO for these trees or in the height of the trees (number of intermediate nodes), we deleted the outlying string.[2] The jackknife procedure resulted in discarding three strings for one subject, two strings for another subject, and one string each for three subjects. One subject had no strings deleted. The resulting jackknifed trees are shown in Figure 3.

Even casual observation of the trees reveals that there is a striking similarity in clusters across subjects. For example, four landmarks—Huron River, the train station, the Broadway bridge, and the Farmer's Market—frequently appear in the same cluster. Although at no point do we analyze data across subjects, we expected similarity of this sort. To us, the consistency across subjects suggests that the structures uncovered by the ordered tree algorithm reflect actual clustering in the representations that subjects have of this space. The similarity arises because different individuals tend to interact with and perceive an environment somewhat similarly (at least if they have a good deal in common, as do students).

Retesting. There is yet another way to assess whether the obtained cluster organizations reflect lasting representations of the space. We had subjects return 6 weeks after initial testing to generate a new set of recall protocols (one subject had since left school).
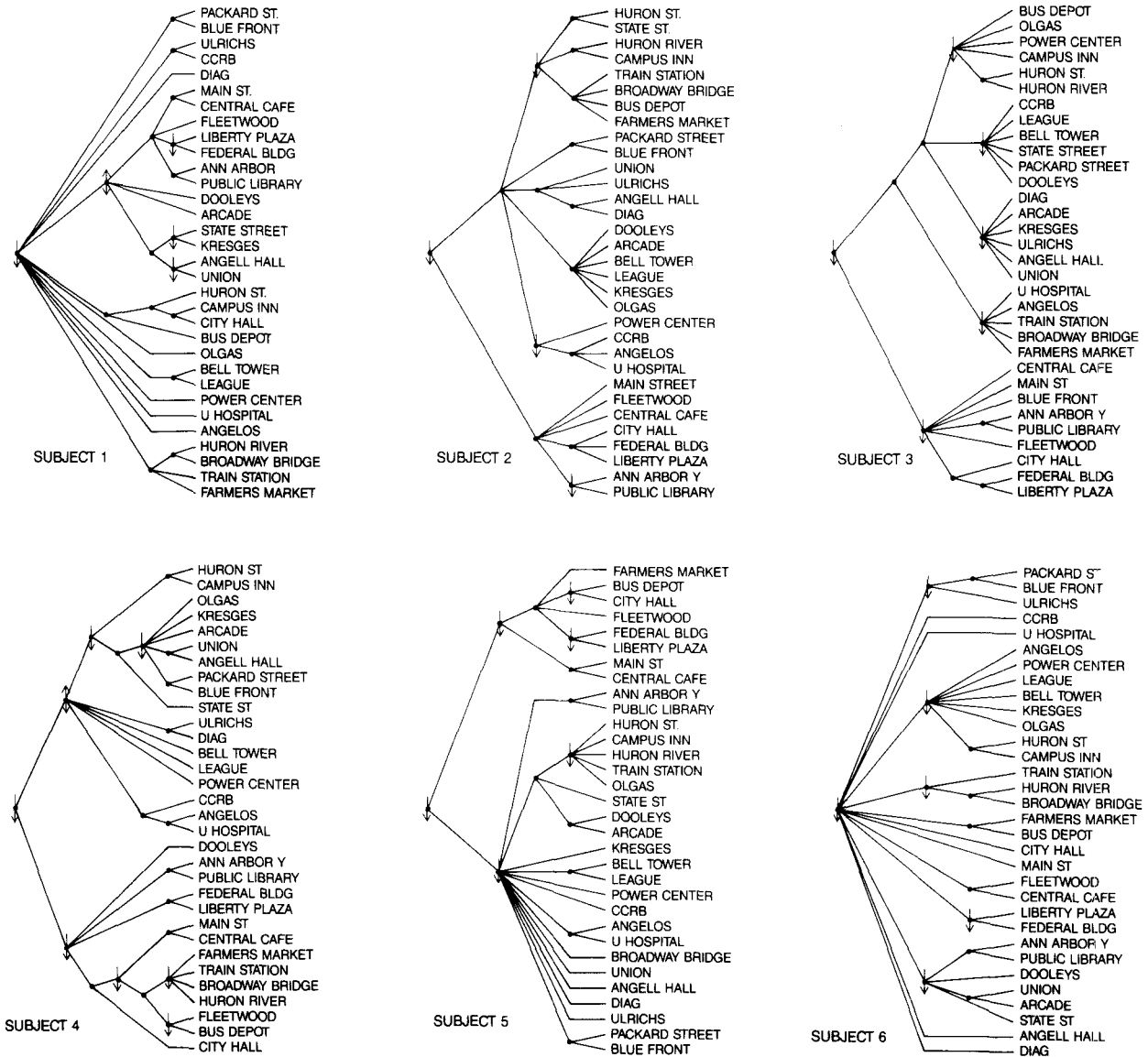
Figure 3. Individual ordered trees derived from the recall data of each subject.

The subjects were first asked to recall as many of the landmarks as possible without being prompted. None of the subjects was completely successful. The average number of items recalled from the set of 32 items was 23.5, with a range of 17 to 28. In three cases, an additional item that was not in the recall set was manufactured by the subject. Overall, then, one can see that the recall set was lost from memory, at least partially.

The subjects were then shown the 3×5 stimulus cards, given a brief time to memorize the names once again, and tested with an identical set of cued and uncued trials, as before. Similarity of the original and replicating trees was quantified using a measure developed specifically for ordered trees (Hirtle, 1982; McKeithen et al., 1981). The measure is defined by:

$$sim(S,T) = \frac{\ln(\mid R_X \cap R_T \mid)}{\ln(\mid R_S \cup R_T \mid)}$$

where S and T are two trees, R_X is the set of chunks in tree X, and $\mid \mid$ denotes cardinality of the set.

The measure is calculated by dividing the number of chunks in common by the total number of chunks in the two trees. The logarithmic transformation is included because the growth of chunks is not linear. Although there is some difficulty in judging this measure in *absolute* terms, it has been used successfully in a relative manner (McKeithen et al., 1981).

In the present study, the average similarity across subjects between the original and replicating trees was .681

for four of the five subjects. One subject, Subject 2, whose tree is shown in Figure 4, had a particularly low similarity between his original and replicating trees, .296. However, this was due in part to three landmarks: City Hall, State Street, and Ulrich's. Removal of these three items resulted in trees whose similarity was .508.

In order to achieve a better sense of the meaning of these similarity measures, similarities were calculated for all pairs of original and replicating trees without regard to the subject who contributed the tree. With the exception of Subject 2, the similarity from a subject's original tree to his or her replicating tree was greater than it was to any other subject's tree. Thus, for all but one subject, the organization as reflected by the original trees appeared to remain stable across time. For the remaining subject, 29 of the items remained consistent, whereas 3 of the items were apparently grouped differently.

Having established the long-term reliability of the trees, we next sought measures of their validity. This we examined by testing whether the clusters that are indicated in the trees help to predict performance in tasks that require knowledge of the space. Two kinds of tasks were evaluated. One task required knowledge of relative distances in the space. The other set of tasks, of which there were three versions, assessed subjects' knowledge of absolute distance.

Relative Distance Judgments

The critical stimuli for this task consisted of 32 triplets of landmarks; each triplet consisted of two anchors and a target. The triplets were constructed such that the two anchors of the triplet were roughly equal in distance from the target of the triplet. The subjects judged whether a target/anchor pair was short or long relative to a standard. Thus, each subject judged 64 pairs of landmarks

Figure 4. Histogram of error percentages by distance category.

or 32 pairs of pairs of landmarks. Since there were six subjects in the experiment, there were 192 total pairs of pairs of landmarks that were critical to our analysis.

In order to analyze whether clustering had an effect on relative distance judgments, we first had to select from the 192 pairs of pairs those that had one doublet of landmarks in the same cluster and the other doublet in different clusters. Since the 192 pairs of pairs come from all the subjects, this is, naturally, a judgment that must be made for each subject. We used two criteria to make this judgment:

(1) Consider a pair of pairs T_1-A_1 and T_1-A_2. One way in which these landmarks might be related for a subject is indicated by the following (where the brackets indicate any of the three types of chunk delimiters): $\{ \ldots \{ \ldots T_1 \ldots A_1 \ldots \} \ldots A_2 \ldots \}$. In this case, T_1 and A_1 were judged to be a within-cluster pair, whereas T_1 and A_2 were an across-cluster pair. There were 88 such sets of pairs, with one member of the set classified as within cluster and the other classified as across cluster.

(2) Consider a triplet of landmarks of the following form: $[\ldots T_1 \ldots A_1 \ldots A_2 \ldots]$. This represents a unidirectional cluster, with A_1 always recalled after T_1 and A_2 always recalled after A_1. In this case, we judged A_1 to be closer in the mental representation to T_1 than was A_2 to T_1. This classification may be controversial, but in terms of the relative distance judgment analysis, it is conservative, since we are claiming that T_1-A_2 is an across-cluster pair and T_1-A_1 is a within-cluster pair, and we are classifying the respective response times and errors accordingly. If T_1, A_1, and A_2 are, in fact, all members of the same cluster, our procedure is biased against finding a difference between within-cluster and across-cluster pairs. This criterion yielded 34 additional pairs of pairs for analysis. So, of the 192 critical pairs of pairs of landmarks in the experiment, 122 participated in the analysis of relative judgments.

Our definition of pairs as being within versus across cluster is, then, a relative definition. For example, given the structure for Subject 1, we find Huron St.-Campus Inn to be an *across-cluster* pair relative to Campus Inn-City Hall, whereas it is a *within-cluster* pair relative to Huron St.-Bus Depot.[3] This judgment was made individually for each subject. For example, the pairs League-CCRB and League-Campus Inn were classified, respectively, as within and across cluster for Subject 3, and as across and within for Subject 6.

Errors. First, we examined the proportion of errors made by subjects to these 122 pairs of pairs of landmarks. An error is defined as a judgment by a subject that an anchor was close to the target (compared with the standard) when it was, in fact, far, or vice versa. Of the 122 pairs of pairs, 51 actually represented short distances and 71 represented long distances (compared with the standard). Figure 4 displays the proportion of errors as a function of whether a distance was actually short or long, and as a function of whether a pair was within or across
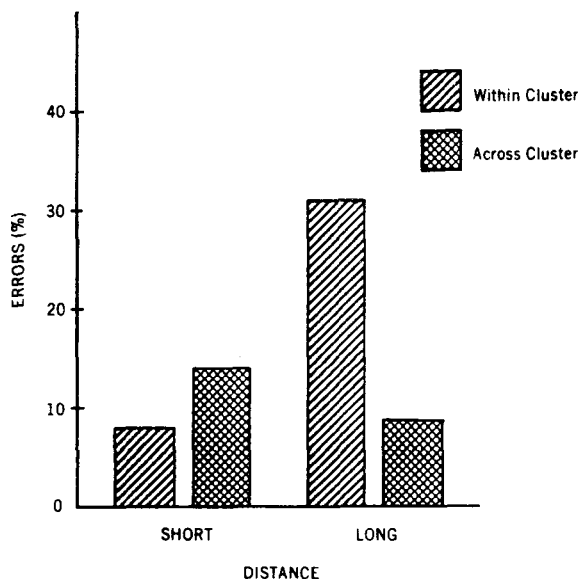
cluster. It is quite clear that there is an interaction in these data, as expected: For short distances, subjects tend to classify a distance as "long" if a pair is across cluster compared with within cluster, and to do quite the opposite for long distances. This interaction is statistically significant [$\chi^2(1) = 6.33$, p < .05]. In addition, we examined response times. However, no significant differences emerged between the times; thus, differences appear only in the error rates, not in the response times.

Absolute Distance Judgments

There were three tasks in which subjects provided data about absolute distance: (1) They explicitly gave magnitude estimates of distance, (2) they constructed maps of the space from block figures, and (3) they drew sketch maps. We analyzed each of these sources of data.

Distance estimates. There were magnitude estimates provided for the 64 critical pairs that were described above. Overall, these estimates were quite accurate, as indicated by the mean correlation (.803) between the estimated and actual distance. This value is not as high as that found in previous studies (Golledge & Zannaras, 1973; Jonides & Baum, 1978), but the range of distances was more restricted in the present study. Here, the ratio of the longest to the shortest distance was 5.2:1. For the aforementioned studies, the ratios of distances were approximately 23:1 and 115:1, respectively.

The pairs were classified as within or across cluster using the method described above for the relative distance task. Then these data were subjected to a stepwise multiple regression analysis including the factors of distance, cluster, and subject. All three factors yielded reliable effects [$F(1,236) = 375.6$, p < .001, for distance; $F(5,236) = 19.69$, p < .001, for subject; $F(1,236) = 7.25$, p < .01, for cluster]. The averaged regression lines for each cluster are displayed in the scatterplot of Figure 5. Regression lines have been added for the within-cluster and across-cluster pairs. Note that only the inter-
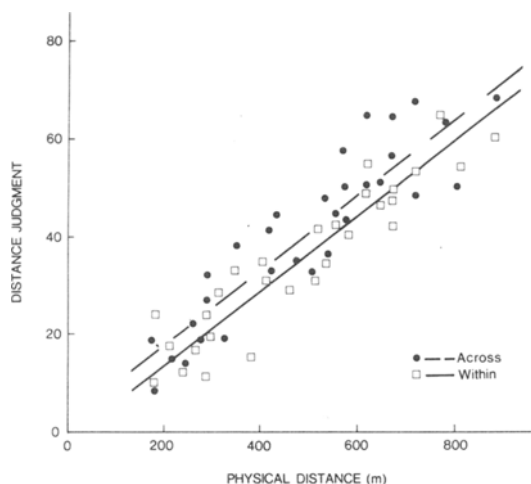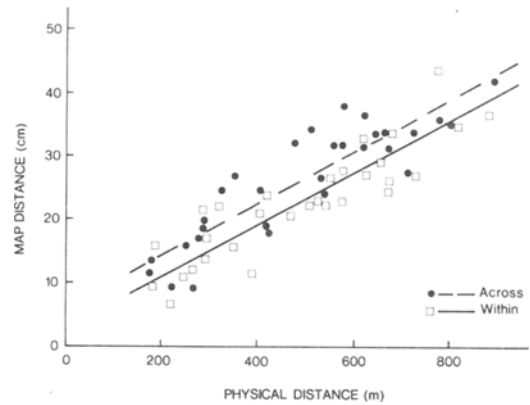

Figure 6. Scatter plot of constructed map distances by physical distances.

cept of the function, not the slope, varies with the cluster variable. These overall results are consistent with those of both Kosslyn et al. (1974) and Maki (1981), although the effect in our data is somewhat small. The change in R-square due to cluster was .01, resulting in an R-square value for the multiple regression of .67. To help verify that the effect was real, we next examined the maps that subjects constructed from blocks.

Constructed maps. We recovered distance estimates from these maps simply by measuring the distances between landmarks for our critical within- and across-cluster pairs. These data are, perhaps, more accurate reflections of the quality of subjects' distance knowledge, since subjects had not been given feedback about distances prior to this phase of the experiment, whereas they had been given feedback prior to the distance-estimation task. For the map data, the average correlation between estimated and actual distance was .725. Once again, we used stepwise multiple regression to analyze these data. All three effects were reliable [$F(1,236) = 248.6$, p < .001, for distance; $F(5,236) = 21.56$, p < .001, for subject; $F(1,236) = 12.93$, p < .001, for cluster]. These data are plotted in Figure 6. The R-square value for all three factors was .61, and the change in R-square for the cluster factor was .03.

Sketched maps. The third source of information about an effect of clusters on distance estimates is the maps that subjects sketched. Although there were four such drawings per subject, we examined only one, the fourth, since it presumably was produced after the subjects had grown accustomed to the 8½ × 11 in. restriction on the drawing surface. The distance estimates from these maps were correlated .627 with the actual distance. Using stepwise multiple regression, we found a similar relationship, with all three predictors again showing reliable effects [$F(1,236) = 139.3$, p < .001, for distance; $F(5,236) = 8.83$, p < .001, for subject; $F(1,236) = 9.98$, p < .01, for cluster]. The overall R-square was .45, and the change in R-square for cluster was .03. These data are plotted in Figure 7.
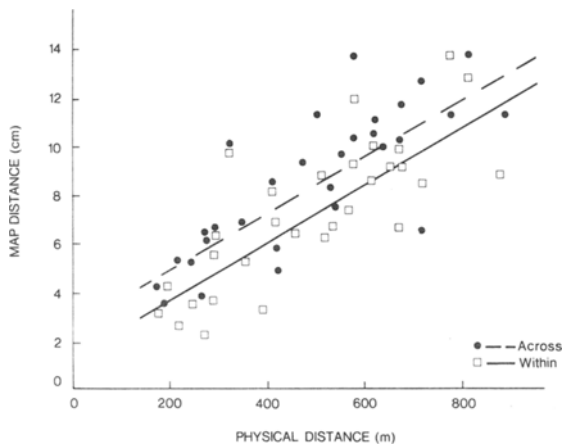

Figure 5. Scatter plot of distance judgments by physical distances.

Figure 7. Scatter plot of drawn map distances by physical distances.

DISCUSSION

We have offered evidence of an effect of subjective cluster membership on performance in several tasks that require information about distances in a natural space. To summarize the results from all the tasks, across-cluster distances tend to be overestimated, whereas within-cluster distances are underestimated, relative to each other.

This summary, of course, is based strictly on our measurement of subjective clusters, a measurement that is intuitively pleasing and, apparently, stable over time. But what if it is basically in error? What if mental representations of real physical spaces are not truly hierarchical. Will the ordered tree algorithm impute a fundamentally incorrect organization to the space?

Yes and no. Consider perhaps the most reasonable alternative to a hierarchical organization, a network such as the one illustrated in Figure 8. What would happen if we were to create an ordered tree structure from recall data that had actually been a result of such a network structure? Given adequate sampling of recall protocols, the resulting tree would be a series of clusters all rooted to a single nondirectional node. So the good news is that the ordered tree algorithm would recover existing clusters. The bad news is that it would impose a hierarchy where none existed. Of course, the imposed hierarchy is really a degenerate case (no subordinate nodes), an outcome that we have not observed and that might have raised our suspicions if we had observed it. Since we made predictions only about clusters and not about intercluster organization for our performance tasks, our use of the ordered tree algorithm appears to be on safe ground. But until a space of alternative possible organizations has been explored, one must remain cautious in making too much of hierarchies such as those pictured in Figure 3. But these hierarchies, taken together with the performance data, certainly invite the conclusion that we have recovered some regularities in the representation of spatial information.

A second concern is that the structure of landmarks is not uncovered, but is induced, by the free-recall task. That is to say, the structure may be a by-product of recall strategies. Later, when subjects reconstruct the space and produce distance judgments, they may be relying on the structure recently imposed to access information, and, in turn, this structure results in the observed biases. Unfortunately, the stability over time does not answer this criticism, because the stability may be due to partial recovery of the original retrieval strategy. The resolution of this conflict centers on two larger issues. First, there is a fundamental issue of what structure the ordered tree algorithm taps. There is converging evidence from several studies (McKeithen et al., 1981; Reitman & Rueter, 1980) that the algorithm does in fact tap existing long-term memory structures. Second, even if the structure was induced (which is unlikely), our data still indicate that nonspatial information influences spatial judgment. Thus, the data show that hierarchies can lead to bias in spatial judgments. The exact nature of how hierarchies arise and the strength of the bias need further research.

Let us examine the claim that subjective clusters influence knowledge about spatial relationships. To be sure, we are not willing to conclude that hierarchies of the sort that we have illustrated are a sufficient representation of spatial knowledge. They include no explicit information on actual spatial relationships (e.g., distance and bearing) about which subjects clearly have knowledge. Thus, the subjective clusters constitute but one piece of information in a complex data base that subjects must have available for spatial information.

How could subjective clusters bias judgments about distance, as our data demonstrate that they do? Two possibilities suggest themselves. One is that cluster knowledge causes subjects to mentally misplace landmarks by causing an exaggeration of intercluster distance (Allen, 1981; Kosslyn et al., 1974; Maki, 1981). According to this alternative, mistakes in distance judgments can be traced to fundamental misconceptions about the relative locations of landmarks, misconceptions that in turn are caused by cognitive knowledge of those landmarks. If this were the
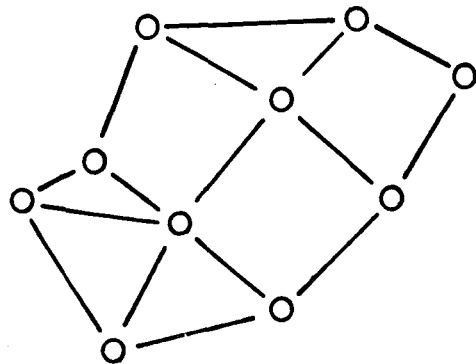

Figure 8. Sample network structure. Circles denote cluster of landmarks; lines denote links between clusters.

case, the clustering of landmarks would be evident in the perceived spatial arrangement and the hierarchies that we uncovered would be a by-product of how subjects name locations off their cognitive maps.

A second possibility is that cluster knowledge has its effect during the process of extracting spatial information about landmarks. This may account not only for the distance judgment data, but also for Stevens's (1976) response time data. It may be, for instance, that cluster organization has an effect on accessing the locations before distance judgments can be made; thus, locations that cross cluster boundaries would be difficult to visualize simultaneously. This interpretation implies that landmarks are not mislocated but rather, due to the mental effort of traversing a boundary, the points are perceived as more distant. If the mental effort could be minimized, then the perceived distance would be less. At present, however, there are no convincing data to resolve the issue of how clusters have their impact.

To summarize, then, our data support the view that mental representations of actual spaces are composed of both spatial and nonspatial information. Furthermore, the nonspatial information is hierarchical in nature even for spaces in which there are no predefined hierarchies. Taken together with previous data (e.g., Sadalla et al., 1979; Stevens & Coupe, 1978), our results force an alteration of a strict mental map model of spatial knowledge. At the least, such a model will have to provide for the representation of non-Euclidean information.

REFERENCES

ALLEN, G. L. (1981). A developmental perspective on the effects of "subdividing" macrospatial experience. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 120-132.

BAIRD, J. C. (1979). Studies of the cognitive representation of spatial relations. *Journal of Experimental Psychology: General*, 108, 90-106.

BRIGGS, R. (1973). Urban cognitive distance. In R. M. Downs & D. Stea (Eds.), *Image and environment* (pp. 361-388). Chicago: Aldine.

CADWALLADER, M. (1979). Problems in cognitive distance: Implications for cognitive mapping. *Environment and Behavior*, 11, 559-576.

CANTER, D., & TAGG, S. (1975). Distance estimation in cities. *Environment and Behavior*, 7, 59-80.

COHEN, R., BALDWIN, L. M., & SHERMAN, R. C. (1978). Cognitive maps of a naturalistic setting. *Child Development*, 49, 1216-1218.

GOLLEDGE, R., & ZANNARAS, G. (1973). Cognitive approaches to the analysis of human spatial behavior. In W. Ittelson (Ed.), *Environment and cognition* (pp. 59-94). New York: Seminar Press.

HIRTLE, S. C. (1982). Lattice-based similarity measures between ordered trees. *Journal of Mathematical Psychology*, 25, 206-225.

HIRTLE, S. C. (1985). *Cognitive structures in cognitive maps: Evidence of hierarchies in spatial representation*. Manuscript submitted for publication.

HOLYOAK, K. J., & MAH, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, 14, 328-352.

JOHNSON, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32, 241-254.

JONIDES, J., & BAUM, D. R. (1978, May). *Cognitive maps: Analysis of distance estimates*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago.

KOSSLYN, S. M., PICK, H. L., & FARIELLO, G. R. (1974). Cognitive maps in children and men. *Child Development*, 45, 707-716.

MAKI, R. H. (1981). Categorization and distance effects with spatial linear orders. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 15-32.

McKEITHEN, K. B., REITMAN, J. S., RUETER, H. R., & HIRTLE, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13, 307-325.

MOSTELLER, F., & TUKEY, J. W. (1977). *Data analysis and regression: A second course in statistics*. New York: Addison-Wesley.

NEWCOMBE, N., & LIBEN, L. S. (1982). Barrier effects in the cognitive maps of children and adults. *Journal of Experimental Child Psychology*, 34, 46-58.

REITMAN, J. S., & RUETER, H. R. (1980). Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology*, 12, 554-581.

SADALLA, E. K., BURROUGHS, W. J., & STAPLIN, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 516-528.

SADALLA, E. K., & MAGEL, S. G. (1980). The perception of traversed distance. *Environment and Behavior*, 12, 65-79.

SADALLA, E. K., & STAPLIN, L. J. (1980a). An information storage model for distance cognition. *Environment and Behavior*, 12, 183-193.

SADALLA, E. K., & STAPLIN, L. J. (1980b). The perception of traversed distance: Intersections. *Environment and Behavior*, 12, 167-182.

SADALLA, E. K., STAPLIN, L. J., & BURROUGHS, W. J. (1979). Retrieval processes in distance cognition. *Memory & Cognition*, 7, 291-296.

STEVENS, A. (1976). *The role of inference and internal structure in the representation of spatial information*. Unpublished doctoral dissertation, University of California, San Diego, La Jolla, CA.

STEVENS, A., & COUPE, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, 10, 422-437.

THORNDYKE, P. W. (1981). Distance estimation from cognitive maps. *Cognitive Psychology*, 13, 526-550.

WILTON, R. (1979). Knowledge of spatial relations: The specification of the information used in making inferences. *Quarterly Journal of Experimental Psychology*, 31, 133-146.

NOTES

1. Cues were used to encourage variety in the recall orders. The phrase "those that went with it" was included to discourage recall patterns of the form $\{i, 1, 2, i-1, i+1, \ldots, n\}$, in which the same linear order is produced on every trial with the cue inserted at the beginning. No other recall patterns were discouraged, and subjects were encouraged to "mix up" the orders as much as possible.

2. As each recall string is individually deleted in the jackknifing process, one can calculate a mean and standard deviation for n trees, each based on a set of $n-1$ recall strings. Outlier recall strings can then be determined by calculating a confidence interval about the mean. If a score falls outside the interval, the corresponding string is deleted (Mosteller & Tukey, 1977). In building the confidence interval, we adopted a conservative criterion by setting alpha to .001.

3. This quadruplet {Huron St., Campus Inn, City Hall, Bus Depot} does not consist of any critical triads. In fact, Huron St. does not refer to a single location, but an entire street. Instead, this example was chosen to convey the relative nature of the across-cluster/within-cluster distinction, which is independent of the location of the points in the space.