Instructional Effects on Spatial and Temporal Memory for Videotaped Events in a Large-scale Environment

Alex Cuthbert, Christopher Stecker, Inna Aleksandrovsky, Sheryl Ehrlich, Nikunj Oza, Paula Rogers*
`(spatial-cog@cogsci.berkeley.edu)`

Institute for Cognitive Studies
University of California, Berkeley
608 Barrows Hall M/S 3020, Berkeley, CA 94720

Abstract

The separability of spatial and sequential mental representations was examined through the use of sketch-maps and ordered event-lists generated by subjects following the viewing of a videotape depicting movement through a natural space. Prior to viewing, subjects were instructed that they would either a) draw a map of the region depicted and place events on the map (map group), b) make a list of the events they saw in the order they saw them (list group), or c) answer some unspecified set of questions following the video (control group). In fact, subjects did all of the above. Although most measures of spatial and sequential accuracy were unaffected by the instructional manipulation, subjects who expected to draw maps were more likely to correctly indicate that the camera had negotiated the space in a figure-eight path, while subjects in the other groups predominantly indicated circular path shapes. None of our analyses provide any strong evidence that an independent spatial representation exists prior to map-drawing. In fact, the similarity between groups suggests that all subjects utilized similar encoding strategies, but that map subjects specifically attended to features of the film which constrain the overall layout of the space. This research raises specific questions about the mechanisms which allow path segments to be integrated into coherent spatial reference frames.

Introduction

The goal of the present research is to understand the spatial and temporal structure of people's internal representations of naturalistic events in a large-scale environment. We studied the effects of instructional manipulations on spatial and temporal memory for a sequence of videotaped events. One group of subjects was told that they would draw a map of the area shown in the video, another that they would list the events they saw in order, and a third that they would simply answer some questions. After viewing the videotape, all subjects performed both a map-drawing and an event-listing task. We were particularly interested in the degree of independence between the two kinds of information and the possibility of trade-offs: would better spatial memory come at the expense of worse temporal memory (and vice versa), or would both be fully and automatically encoded as long as the subject viewed the tape with the expectation of having to remember something about it?

We studied memory for videotaped event sequences as a compromise between considerations of ecological validity and stimulus control. The static spatial tasks often used in spatial cognition studies, such as memory for maps (Liben & Downs, 1991; Thorndyke & Golding, 1981; Cohen & Schuepfer, 1980) and descriptions of computer images (Hayward & Tarr, 1995) afford good stimulus control but have restricted ecological validity and weak generality relative to spatial cognition in everyday life. Naturalistic situations, such as studies of way-finding and direction-giving (Morse, 1987; Chase & Chi, 1981; Lynch, 1960), afford greater ecological validity and generality, but at the cost of uncertain conclusions due to uncontrolled stimulus variables. Our videotape was filmed in a large-scale environment that contained both man-made elements (buildings, roads, etc.) and natural elements (trees, bushes, etc.) as salient landmarks. Viewers saw strategically placed characters, each engaged in an activity such as juggling, clowning, jumping rope, etc. The camera traversed a figure-eight path twice through the space. Some of the events occurred at the same location both times around, and others changed their location between the first and the second pass. This design was employed to allow spatial and temporal aspects of subjects' memory performance to be dissociated.

The theoretical position developed by Siegel & White (1975) to explain the construction of spatial representation posits a stage-like model where people acquire more abstract, procedural knowledge from, declarative knowledge of specific landmarks and routes. Specifically, they proposed that people first remember landmarks and organize them sequentially into knowledge of routes. With further experience, landmarks and routes are partially coordinated and only later integrated into map-like, global configurations (Millar, 1994). However, additional evidence exists that several factors can influence how spatial information is encoded, including semantic and physical features of the map (Holding,1994) and the goals of

---

*Authors names are listed alphabetically in two groups: Chris Stecker and Alex Cuthbert were primarily responsible for the text, while the design, execution, analysis, and interpretation are jointly due to a collaborative effort involving all authors equally. Requests for reprints or further information should be addressed to Alex Cuthbert, `alx@socrates.berkley.edu`.

the learner (Magliano, 1995; Taylor & Tverksy, 1992; Cohen & Cohen, 1982; Gauvain and Rogoff, 1986).

Magliano et al. (1995) tested the stage-based model of Seigel and White (1975) against a hypothesis that instructional emphasis on a particular level of space encoding (i.e., landmarks, routes, or global configuration) can disrupt the hierarchical learning. In its strong form, this hypothesis postulates that space encoding is goal-constrained and does not conform to the strict hierarchy from individual landmarks through routes to global overview. Their findings indicated that subjects were indeed capable of learning a new environment according to a goal, but learning was constrained by stage-based processes. In other words, a hierarchical local-to-global structure underlies space encoding, but could be to some degree modified by the instructions, providing evidence for a dissociation between encoding stages.

In a similar vein, we set out to determine if a similar dissociation can be achieved between spatial and sequential information encoded during naturalistic viewing of a space. We hypothesized that subjects expecting to draw maps after viewing would demonstrate enhanced spatial memory, while subjects expecting to generate ordered lists of events would show better sequential memory, at a possible cost of disrupted spatial memory, and that this differential biasing towards spatial or sequential coding would indicate separable mechanisms for spatial and sequential encoding.

Methods

Subjects

The subjects were 39 undergraduate students from the Psychology subject pool at U.C. Berkeley who volunteered to participate in the study in order to fulfill a course requirement.

Materials

The subjects viewed a videotape of approximately ten minutes. The resulting data set included a) ordered lists of icons representing characters in the film, b) drawings of the region seen in the film, c) a set of characters placed on the maps, d) traces of the perceived path through the space, e) responses to imagistic and verbal memory tasks, and f) questionnaires about navigational capabilities.

Each video consisted of two passes along a figure-8 path through a space that consisted of natural and man-made features. From the first to second pass through the space, half of the events moved locations. The film condition involved two films where the moving versus stationary events were interchanged in an attempt to control for the salience of different events. The films included moving versus stationary events to permit assessment of the effect of task focus (map versus list) on the representation of events linked to or independent of a specific location

Figure 1 provides a schematic representation of the area along with the position of the events. The numbered circles stand for the position and sequence of different events. The events that end in an "A" or "B" represent moving events. The events that end in an "A" appear during the first pass through the figure-8 and those ending in "B" appear during the second pass. The other events are stationary events.

The space was separated by a road with a large building on one side of the street and a park on the other. We purposely chose a combination of man-made and naturalistic settings so as not to bias the results in favor of subjects that had a disposition to structured versus unstructured or man-made/naturalistic settings.

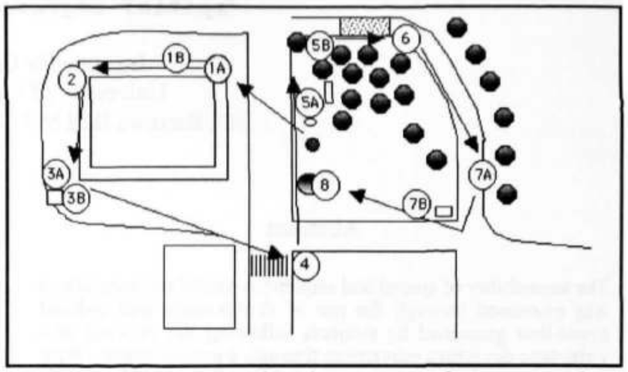

Figure 1: Layout of region, events, & camera path.

Design and Procedure

A 3x2x2 factorial design was used with cells representing instructional condition, task-sequence (list-map versus map-list) condition, and film condition. The three instructional groups (map, list, or control) are determined by the written instructions given prior to viewing the video. The instructions for the map condition state that subjects be asked to draw a map of the space depicted in the film. The instructions for the list condition state that they will be asked to make a list of events depicted in the order they appear in the film. The control condition received instructions stating that they would be asked some questions following the film. The task-sequence condition is based on the order in which the tasks are completed following the video (i.e., list-map versus map-list). This control was included to deal with possible order effects of the tasks upon each other.

Experimental sessions lasted one-hour. The subjects were seated in front of a video monitor where they received one of three instruction sheets. After reading it, they viewed the ten-minute video that included eight salient characters (a clown, a juggler, etc.) each of whom appeared twice in the film. The subjects completed several tasks following the video including a) ordering icons representing characters in the film in the order they appeared (the list task), b) drawing the region seen in the film and tracing the path through which the camera traveled (the map-drawing task), c) placing icons representing characters in the film on the map, d) performing imagistic and verbal memory tasks, and e) answering questionnaires about navigational capabilities. Two copies of the icons representing the characters are provided for both the list task (item a) and the map task (items b and c).

Half of the groups, split uniformly across instructional and film segment conditions, completed the list sequence task followed by the map sequence task or vice versa to control for priming effects.

The image memory task consisted of showing nine images on an overhead projector for thirty seconds. Following a distractor task, subjects had to determine which images they had seen from a larger set of images. This larger set included similar images to the ones they had seen but depicted from different perspectives. Subjects were scored on the number of correctly identified images minus the number of incorrectly identified images. The word memory task consisted of thirty words read aloud by an experimenter at five second intervals, followed by a test of free recall.

Data Coding & Analyses The topological structure of camera path depicted by subjects (i.e., line, figure-8, circle, other) was coded as an outcome measure along with a metric for the temporal and spatial ordering tasks. The number of landmarks present in the map was also analyzed.

In order to examine the degree of separation between subjects' spatial and sequential representations, two comparable dependent measures were generated from the ordered event lists and the hand-drawn maps. First, the *temporal sequence* score reflects the number of transpositions necessary to transform a subject's event list into the actual sequence of events that appears in the film. Second, the path indicated on each subject's map was followed, beginning at the indicated starting point, and each event was listed, in the order it appeared along the path, to generate a *map sequence*. This sequence, in turn, was converted into a score indicating the number of transpositions necessary to transform the map sequence into the correct map sequence, as would have been generated by following the same procedure using a veridical map of the space.

The sequencing tasks were scored by the number of transpositions needed to convert the given sequence to the correct sequence. They were calculated using a sorting algorithm that we devised such that:

correct:	1, 2, 3, 4, 5
subject A:	2, 1, 3, 4, 5
subject B:	4, 1, 2, 3, 5

would produce values of "1" for both subject A and subject B. Subject A had one transposition while subject B had one shifted event. If we merely counted the number wrong, the score would be 2 for subject A and 4 for subject B based on number of items not in the correct cells. The sorting algorithm provides a better indication of performance because it gives credit for shifted but correct sequences of events.

Note that, since half of the events in the film actually changed locations on the second pass through the space, the correct temporal and map sequences differ markedly.

Complete independence of spatial and sequential encoding predicts that the ideal spatial subject (the subject who encodes spatial information perfectly, but ignores sequential information completely) would produce an accurate map sequence, but an inaccurate temporal sequence, with errors due to inability to differentiate items from the two passes through the space. The ideal sequential subject would produce the accurate temporal sequence, with the associated cost of more errors in the map sequence. It was the intention of our primary experimental manipulation to bias subjects to act in one of these two ways.

In addition to temporal and map sequence scores, both an adjacency score (Rovine & Weisman, 1989) and a landmark association score were assigned for each event to permit the calculation of correlations for stationary versus moving events. Following Rovine & Weisman's methodology (1989), we have not scored items that were not present or not explicitly linked to a landmark. Though we know that sketch maps are a good approximation of way-finding skills (Blades, 1990) , it is difficult to assesses mental representations without verbal protocol analyses and records of the sequence of actions (see section on Future Research.) Knowing the order in which events were placed on the map would allow us to say more about the primacy of spatial or temporal representations.

Results & Discussion

The order of presentation of film segments did not have a significant effect ($p > .05$) on any of the dependent variables reported here, nor did the order of experimental tasks (map-drawing and event-listing). With this in mind, we collapsed the data over these two variables for the ensuing analyses. Memory tests of word-list recall and image recognition were included in the design as potential independent predictors of performance. Word recall was not significantly correlated with performance on the temporal sequence ($R^2 = 0.001$, $p > .05$) or map sequence ($R^2 = 0.029$, $p > .05$) performance measures (see below), and so was kept out of the remaining analyses. Similarly, image recognition memory was not found to correlate significantly with either temporal sequence ($R^2 = 0.005$) or map sequence ($R^2 < 0.001$, $p > .05$) accuracy, and was also kept out of later statistical analyses.

Temporal & Map Sequences No significant effect of instructional condition on the accuracy of the temporal sequence was observed ($F_{(2,38)} = 2.76$, $p > .05$), although subjects in the list condition performed highest overall, while control (no instruction) subjects made the most errors (see fig. 2). The effect of instructional condition on map sequence accuracy was also not significant. ($F_{(2,38)} = 0.99$, $p > .05$) Map sequence accuracy displayed the same pattern of results as temporal sequence accuracy (see fig. 3). Since most subjects made few errors overall on these measures, the lack of

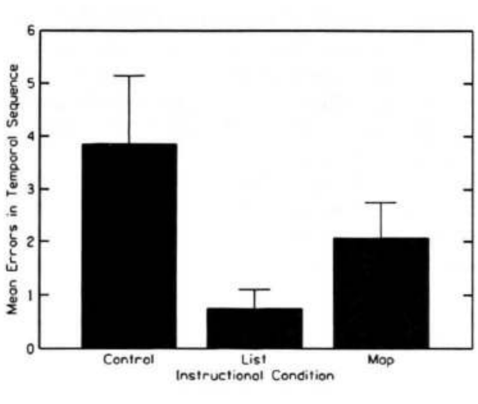

Figure 2: Temporal sequence scores per instructional condition.

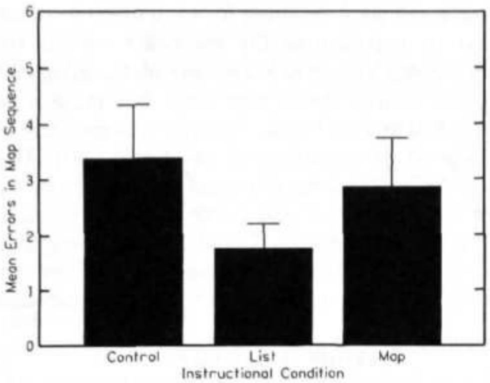

Figure 3: Map sequence scores by instructional condition.

significant differences may be reflective of performance-ceiling effects. However, while trends towards small differences between the instructional conditions were observed, no differential effects due to instructional condition were observed, as would have been seen if sequential or spatial biasing were actually taking place. These results indicate that both the map sequence and temporal sequence measures reflect the results of a single process or representation.

Path Topology One difference that does arise between subjects in different instructional conditions is seen in the distribution of path topologies indicated on subjects' maps (see fig. 5). Most subjects in the map condition indicated the path as traversing a figure-eight, while subjects in the list and control conditions predominantly drew circular paths ($\chi^2_{(1)} = 5.81$*, $p < .05$). This finding indicates that subjects in the map condition were more accurate in reproducing large-scale aspects of the scene layout. Interestingly, this difference is not accompanied by a difference in the number of landmarks displayed on the maps ($F_{(2,38)} = 0.09$, $p > .05$), nor were map subjects more successful at correctly linking events with salient landmarks ($F_{(2,38)} = 0.06$, $p > .05$). In addition, control subjects, who were not instructed as to the type of information to encode, produced path shapes which resembled that of the list subjects, suggesting that control subjects may have, by default, utilized an encoding strategy similar to that of list subjects. This may reflect a bias toward sequential encoding of video sequences, perhaps due to greater difficulty in encoding spatial representations.

The enhanced global spatial accuracy indicated by the large number of figure-eight paths among map subjects raises the question of how that accuracy is achieved. Two possibilities exist for this difference: (1) subjects in the map condition are encoding configurational information separately from sequential information, and are using this configurational information at the time of map-drawing, while list subjects must attempt to construct a spatial representation from the sequential information used for the ordered event listing, which ultimately leads to maps which preserve sequential order but distort the overall geometry of the space; (2) subjectsfocus primarily on sequential information while encoding the film, and construct spatial representations only at map-drawing time. Map subjects specifically attend to aspects of the scene, such as shared landmarks, path-crossings, and metric relationships, which other subjects miss, and that these pieces of information act to critically constrain the overall path topology. For instance, attending specifically to the point at which the path crosses itself from a new direction may provide subjects with the information that the path has a figure-eight shape, without requiring the use of a separate spatial representation.

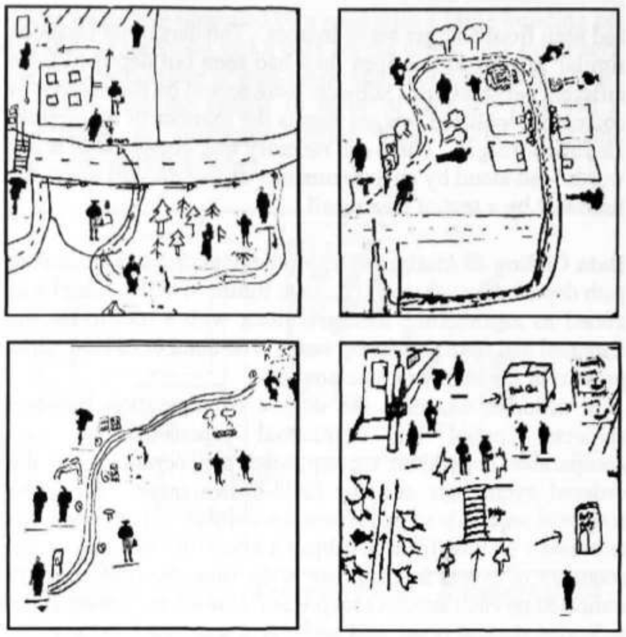

Figure 4: Examples of hand-drawn maps with various path topologies. Clockwise from top left: figure-8, circular, linear, and other.

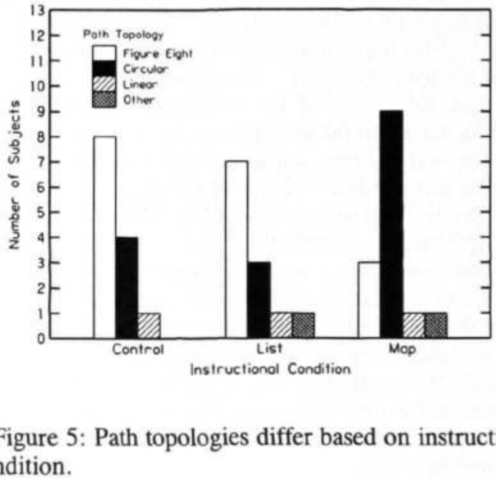

Figure 5: Path topologies differ based on instructional condition.

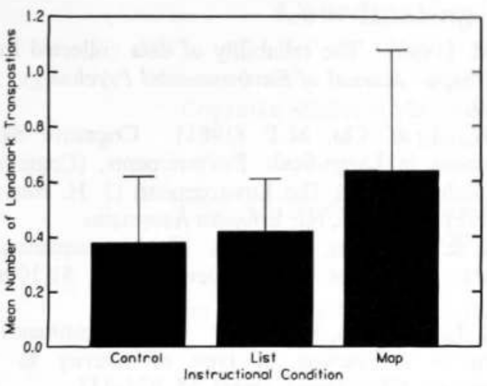

Figure 6: Frequency of landmark transpositions doesn't vary with instructional condition.

Error Analysis: Landmark Transpositions To differentiate between these two explanations of the result, we looked at two other aspects of the hand-drawn maps: errors which simultaneously distort the placement of landmarks and coded events, and differences between adjacency scores for moving and stationary events. Both of these analyses are based upon adjacency scores (Rovine & Weisman, 1989), which indicate whether or not an event has been placed in the correct location relative to its immediate neighbors along the path. In addition, a landmark association score was coded for each event, indicating whether or not a salient landmark is placed near an event to which it is proximal in the actual scene. We used these two measures to look for particular types of errors, *landmark transpositions,* in which an event has been placed incorrectly, in terms of its adjacency score, but is correctly associated with a salient landmark. This indicates that the landmark has been displaced along with the erroneous event, and provides evidence against a representation of the space (including landmarks) which is independent of the event representation. If subjects in the map condition are maintaining such an independent representation, then they should display fewer landmark transpositions than subjects in the list condition, who presumably are not utilizing an independent spatial representation. The results (see fig. 6) indicate no significant difference in the number of landmark transpositions between the groups, ($F_{(2,38)} = 0.20$, $p > .05$) and any trends are in the wrong direction, with map subjects displaying more landmark transpositions than either of the other groups.

Accuracy for Stationary & Moving Events Another measure which might provide evidence for or against map subjects using an independent spatial representation would be a difference in the accurate placement of moving and stationary events. The presence of an independent spatial representation would imply an advantage for stationary events, which appear twice in the same location, over moving events, which appear in different locations on each pass through the space. Map subjects, therefore, should be more accurate (adjacency scores) than list subjects for stationary events, and this difference should be smaller for moving events, which list subjects should represent no differently from stationary events. Our results, however, indicate no significant difference between instructional conditions for adjacency scores of either stationary ($F_{(2,38)} = 0.78$, $p > .05$) or moving events ($F_{(2,38)} = 0.27$, $p > .05$). Non-significant trends exist, in fact, in the wrong direction: map subjects performed worse than list or control subjects for stationary events, but performed slightly better than either list or control subjects for moving events (see fig. 7).

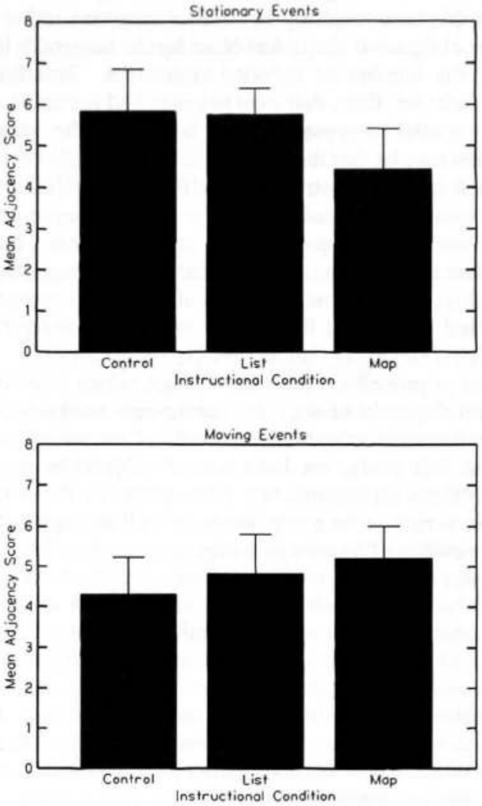

Figure 7: Adjacency scores don't differ between instructional conditions.

Summary & Conclusions

One somewhat surprising finding of this study is the ability of subjects to develop global spatial knowledge of a region from brief views of an area presented through the two-dimensional video display. Independent of condition, subjects were able to pick up information about the structure and layout of the environment, implying that space is either encoded incidentally, along with sequential information, or that it can be reconstructed from sequential knowledge with a fairly high degree of accuracy.

Taken together, our findings seem to indicate the lack of an independent spatial representation in subjects that were aware, during film viewing, that they would be required to construct maps of the depicted space. The finding that map subjects were more likely to correctly draw the camera's path in a figure-eight layout, while list and control subjects predominantly drew circular paths is interesting and somewhat unexpected. This finding seems contradictory to the commonly held view that survey knowledge emerges from route knowledge, and that this process is mediated by landmarks. Instead, focusing on the sequence of events seems to interfere with the higher-level structural representation of space: in this case resulting in circular diagrams rather than the veridical figure-8 shape traversed by the camera, without affecting the number of encoded landmarks. This finding seems to indicate, then, that map subjects had access to more accurate spatial representations; however, the simplest explanation may be that they were able to notice aspects of the film which helped constrain their efforts at constructing an accurate spatial representation at the time of map-drawing. None of our analyses provide any strong evidence that an independent spatial representation exists prior to map-drawing in our subjects. One interpretation of these findings can be made based on current theories of route and configuration knowledge (Siegel & White, 1975; Millar, 1994). The integration of procedural route knowledge, which is necessary for the development of accurate configurational knowledge, normally occurs over extended periods of interaction with a space. In this study, we have forced subjects to generate configurational representations of a space, in the form of hand-drawn maps, after a very short, as well as degraded (i,e. passive viewing of videotape), exposure. It may be that all subjects are compiling route knowledge based on what they see, and that the procedural nature of this route knowledge enables subjects to accurately relate information about the film in terms of sequence, adjacency, and local relationships between events and landmarks. Modifying the goals of the map subjects by informing them of the upcoming map-drawing task causes them to specifically attend to aspects of the scene which would normally (after prolonged exposure) be used to integrate information from several independent routes into a configurational representation. The premature integration of configurational knowledge is assisted by the extra attentional processing applied to the critical junction-points of the routes.

Future Research Clearly, a more careful analysis of the tasks described here can be accomplished by measuring dynamic aspects of map-drawing, through the use of video protocols. Features such as the timing and grouping of map features may provide better evidence for or against independent spatial representations. In addition, planned linguistic and gestural analyses of elicited descriptions of videotaped scenes will provide a slightly different look at the mental representations of space.

Acknowledgements

This research has been a cooperative effort of past and present recipients of the spatial cognition training grant at U.C. Berkeley. We would like to acknowledge the contributions of Steve Palmer, Dan Slobin, and previous members of this group: Christine Diehl, Jonathan Segal, and Nancy Urban. This work was supported by NSF 9355034.

References

Blades, M. (1990). The reliability of data collected from sketch maps. *Journal of Environmental Psychology*, 10, 327-339.

Chase, W. G. &. Chi, M.T. (1981). Cognitive Skill: Implications in Large-Scale Environments, (Cognition, Social Behavior, and The Environment) (J. H. Harvey, Series Ed.). Hillsdale, NJ: Erlbaum Associates.

Cohen, R. &. Schuepfer, T. (1980). The representation of landmarks and routes. *Child Development,* 51(10065-1071).

Cohen, S. L. & Cohen, R. (1982). Distance estimates of children as a function of type of activity in the environment. *Child Development*, 53, 834-837.

Gauvain, M. & Rogoff, B. (1986). Influence of the goal on children's exploration and memory of large-scale space. *Developmental Psychology*, 22, 72-77.

Hayward, W. G. &. Tarr, M.J. (1995). Spatial language and spatial representation. *Cognition*, 55, 39-48.

Holding, C.S. (1994). Further evidence for the hierarchical representation of spatial information. *Journal of Environmental Psychology*, 14, 137-147.

Liben, L. S. &. Downs, R.M. (1991). Developing map concepts in children and psychologists: Going beyond maps as RE-presentations. Paper presented at the *Biennial Meeting of the Society for Research in Child Development*, Seattle, WA.

Lynch, K. (1960). The Image Of The City. Cambridge, MA: MIT Press.

Magliano, J. P., Cohen, R., Allen, G. L., Rodrique, J. R. (1995). The immpact of a wayfinders goal on learning a new environment: Different types of spatial knowledge as goals. *Journal of Environmental Psychology*. 15, 65-75.

Millar, S. (1994). Understanding And Representing Space: Theory And Evidence From Studies With Blind And Sighted Children. Oxford: Clarendon Press.

Morse, J. R. (1987). The construction of perspectives: Piaget's alternative to spatial egocentrism. *International Journal of Behavioral Development*, 10(3), 263-279.

Rovine, M. J. & Weisman, G. D. (1989). Sketch-map variables as predictors of way-finding performance. *Journal of Environmental Psychology*, 9, 217-232.

Siegel, A. & White, S. H. (1975). The development of spatial representation of large-scale environments. In H. W. Reese (Ed.) *Advances in Child Development & Behavior*, Vol. 10. London: Academic Press.

Taylor, H. A., & Tversky, B. (1992). Descriptions and depictions of environments. *Memory and Cognition*, 20, 483-496.

Thorndyke, P. W. &. Golding, S.E. (1981). Ability differences and cognitive mapping skill. Army Research Institute for the Behavioral and Social Sciences.