Spatial Cognition and Computation **1:** 131–144, 1999.

Learning relative directions between landmarks in a desktop virtual environment

WILLIAM S. ALBERT[1,2], RONALD A. RENSINK[2] and JACK M. BEUSMANS[2]
[1]*Department of Geography, Boston University, Boston, MA, U.S.A.;* [2]*Cambridge Basic Research, Nissan Research and Development, Inc., Cambridge, MA, U.S.A.*

Abstract. This study presents two experiments that examine how individuals learn relative directions between landmarks in a desktop virtual environment. Subjects were presented snapshot images of different virtual environments containing distinguishing landmarks and a road network. Following the presentation of each virtual environment, subjects were given a relative direction test. The relative direction test involved indicating the direction of hidden landmarks from different vantage points in the environment. Half of these vantage points were presented during the learning phase, while the other half were novel. Results showed that subjects learned relative directions between landmarks equally well when scenes were presented in either a sequential or random order. Furthermore, viewing a configuration of landmarks in a desktop virtual environment from multiple perspectives produced a viewpoint dependent representation in memory. Subjects had significantly greater response times for new viewing perspectives, as compared to previously viewed scenes. This viewpoint dependent representation of the environment persisted despite learning under conditions of spatio-temporal discontinuity and changes to an environmental frame of reference.

Key words: direction knowledge, frames of reference, spatial cognition, viewpoint-dependence, virtual environments

Introduction

Cognitive mapping research has traditionally focused on how humans navigate and acquire spatial information about real environments, such as cities or neighborhoods (Couclelis, Golledge, Gale and Tobler 1987; Gärling and Golledge 1993; Lynch 1960). However, only recently researchers have investigated how individuals learn to navigate through virtual environments and mentally represent those environments (Albert, Reinitz, Beusmans and Gopal 1999; Darkin, Allard and Achille 1998; Richardson, Montello and Hegarty 1999; Tlauka and Wilson 1996). While using virtual environments to examine cognitive mapping may lack a certain degree of ecological validity, it also offers new opportunities to the cognitive mapping researcher, such as control over layout of the environment and what features in the environment can be seen at any given time.

In the present study, we leverage these advantages by examining how individual's learn relative directions between landmarks in a desktop virtual environment under different viewing conditions and using different environmental layouts. Specifically we are interested in how the presentation order of views (either spatially continuous or discontinuous), the amount of exposure time to the environment, and the presence or absence of an environmental frame of reference all effect learning relative directions between landmarks in a virtual environment. Furthermore, we are interested in the degree to which mental representations of the virtual environments are viewpoint-dependent or viewpoint-independent. These two types of representations are examined by comparing relative direction estimations on old and novel views of the environment.

Presentation order of views has been found to affect the learning of spatial relationships along a route (Allen 1988; Allen, Siegel and Rosinski 1978). Allen (1988) showed subjects a series of slides depicting a walk through an urban environment in either a sequential or random order. In the random condition, subjects' estimates of distances along the route were not as accurate, although their accuracy improved after a second viewing approaching that of the subjects who had viewed the series of slides once in a sequential order. We used the same manipulation but for a series of views from a circular path around a single area with a small number of landmarks. Since Allen's (1988) work had shown that presentation time is important, we also compared the performance of subjects who had unlimited time to study individual views with that of subjects who only had a limited amount of time.

Certain features of an environment may facilitate the representation of spatial relationships among landmarks by providing a frame of reference or by giving directionality to an environment. For example, Subbiah, Veltri, Liu, and Pentland (1996) found that a line of trees in a virtual world improved the ability of subjects to recall the shape of a circuitous route they had had to navigate several times through an otherwise barren environment. The line of trees may have acted as an axis for coding the orientation of each road segment as towards, away from, or parallel to the tree line. In the present study we test the importance of a frame of reference (in the form of a line of trees) within a desktop virtual environment.

Viewpoint dependent and viewpoint independent representations have recently been examined in the context of inter-object spatial relations (McNamara and Diwadkar 1996; Diwadkar and McNamara 1997; Shelton and McNamara 1997). Shelton and McNamara (1997) found that subjects who studied the spatial layout of a collection of objects in a room from two different viewing perspectives (separated by 90 degrees) acquired two viewpoint dependent representations in memory. This was evident by significantly smaller angular error and response latency for previously viewed headings, as compared to the other headings. Diwadkar and McNamara (1997) also observed similar results. Subject studied the locations of various objects on a desktop from a single view, followed by three training views. Results showed that response latency was a linear function of the angular distance to the closest study or training view. Taken together, these results suggest that inter-object spatial relations are represented in memory in a viewpoint dependent manner. Therefore, a novel view of an old scene must be normalized to the closest view in memory, as measured by the smallest angular difference between the study view and new view. The larger the angular difference, the more difficult it is to normalize the new views, as indicated by the linear function in response latencies. In the present study we will examine these findings within the context of a desktop virtual environment.

The virtual environment we test is a small geographic area whose salient features can be seen from a single or a small number of viewing perspectives (comparable to Montello's (1992) "vista space" which describes the area visible from a single vantage point). Each environment contains a unique, moderately complex road network and four buildings on a textured ground plane; building's which are asymmetric and unique in color, shape, and size. After learning an area's layout, subjects were queried with scenes showing only a few of the buildings; subjects had to decide as quickly as possible whether a hidden building was located to the left or right. We compare performance after learning periods using: (i) scenes with or without a frame of reference in the form of a line of trees, (ii) views of the environment presented in a random or sequential order, and (iii) using unlimited or limited time per view. In addition, we compare performance on two types of test views: 'old' views (previously seen by the subject during the learning condition), and 'novel' views (not seen during the learning condition). As a control condition, we also assessed relative direction knowledge after study of the area from only a single vantage point. The results of this study will shed light on how individuals learn relative directions in a desktop virtual environment and the role specific environmental features and viewing conditions play in this learning process.

Experiment 1: Multiple study views

Subjects

A total of 32 (17 female and 15 male) subjects participated in Experiment 1. Subjects were recruited from the Cambridge Basic Research subject pool and were compensated for their participation. Subjects were run in individual sessions lasting one and a half hours.

Stimuli

Four virtual environments were created using the graphics programming language of a Silicon Graphics Indigo workstation. Each virtual environment contained a unique, moderately complex road network and four buildings on a textured ground plane. Each building was unique in color, shape, and size; dimensions of buildings and roads were realistic with a three story-building being 16 m high. The same four buildings were used in all environments; this was done intentionally to ensure that the environments only differed in the spatial configuration of their landmarks. A long row of trees along one edge of the environment was added to provide a frame of reference. Each environment was presented using a series of snapshots, taken from an imaginary circle around the area and facing its center (212.5 m radius; simulated eye height of 4.5 m). Successive snapshots differed by 15 degrees in viewing direction, so that a total of 24 snapshots were used to complete a walk around an environment. The number of buildings in each snapshot varied with its vantage point and the location of the buildings in the environments. Snapshots were displayed on a 20" color monitor with a resolution of 1280 (horizontal) $\times$ 1024 (vertical) pixels. Subjects sat 65 cm from the monitor, so the display subtended 24 degrees vertically and 41 degrees horizontally.

Procedures

Subjects performed a total of eight tests (in two blocks of four tests each); the two blocks were separated by a 30-minute break during which subjects took some general cognitive abilities tests. Half the subjects had the row of trees included during the first block only, the other half had the trees in the second block only. Presentation order alternated between sequential and random; half the subjects started with a sequential presentation order in the first block and a random presentation order in the second block, the other half used the reverse. The starting perspective in the sequential presentation order was randomized to prevent any potential biases in the orientation of the subject's mental representation. Note that the four environments used in the first block were shown again in the second block; this was done to ensure that the effects of trees and scene presentation order (random and sequential) could be compared for the same environment. Since the same four buildings were used throughout, it was not at all obvious that the same configurations were used in the second block and none of the subjects was aware of it as verified during debriefing.

Subjects were instructed to learn the layout of the environment, paying particular attention to the spatial relationships among the buildings. Half of the subjects (n = 16) were allowed to spend as much time viewing each snapshot as they felt they needed to understand the layout of the environment (self pace). The time spent looking at each snapshot was measured as exposure time (ET). The other half of the subjects (n = 16) were presented each of the 24 snapshots for exactly three seconds (fixed pace).

Immediately after seeing the 24 snapshots of an environment, subjects were tested for relative direction knowledge. Subjects were presented a picture of a building from the environment for three seconds (Figure 1), followed by a scene of the environment (Figure 2). Subjects then indicated by a mouse button press whether that building was located towards the left or right with respect to the scene they were viewing. During the direction estimation, a small picture of the building was displayed directly above the scene to prevent possible memory errors (Figure 2). Subjects were instructed to respond as fast as possible, without sacrificing accuracy. Nevertheless, some subjects took extremely long to respond (occasionally timing out at 60 seconds); outliers, defined as response times more than 2 standard deviations above the mean (resulting in threshold response time of 12.7 sec), were excluded from the analysis. This resulted in the exclusion of 174 responses out of a total of 4,080 (4.26%). Test scenes were either the same snapshots as in the learning phase (old views), or novel views of the same environment. Novel views were constructed by adopting a novel vantage point (either slightly closer or further away from the center of the environment) and a unique viewing direction (not oriented towards the center of the environment). A total of 16 test scenes were presented for each environment; 8 were old and 8 were novel. See Figure 3 for the locations and viewing directions of the novel and old vantage points in one environment. The novel vantage points differed for the four environments because of differences in layout. Although old test views may be familiar, they do not contain all the information for a correct response. The presentation order of test scenes was randomized. No feedback on accuracy was given to subjects during the test. The presentation order of each environment and learning condition was counter-balanced.

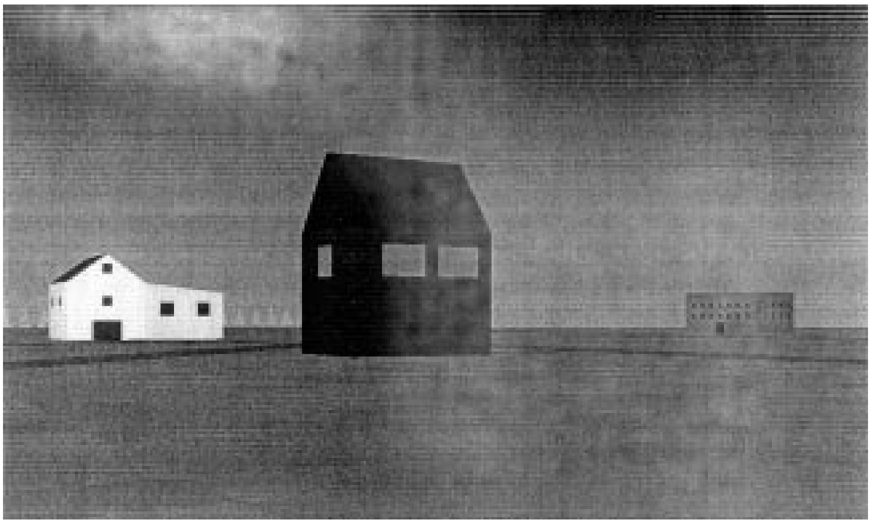

Figure 1. Sample scene during learning phase with trees visible in the background.

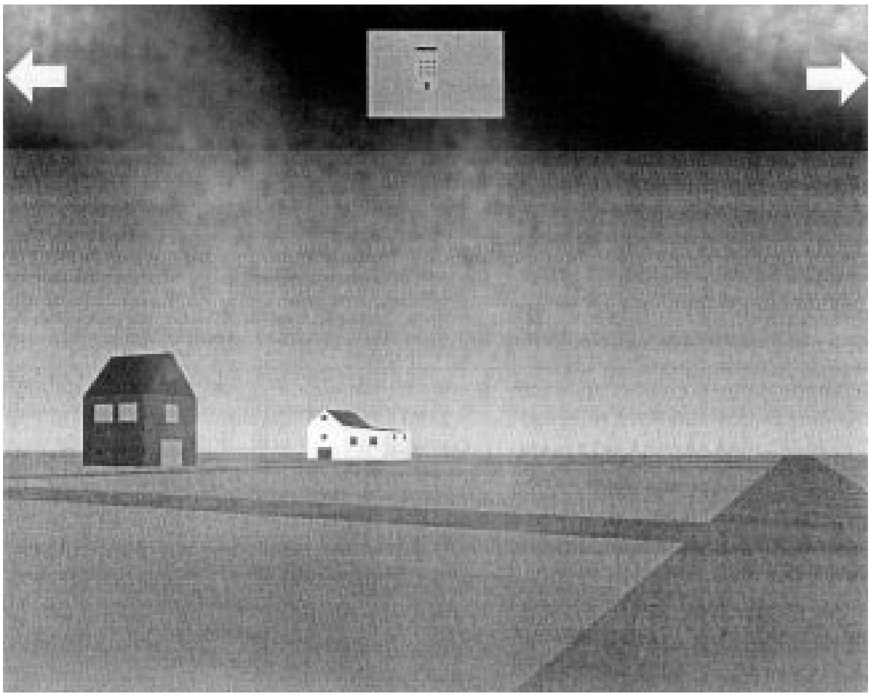

Figure 2. Sample test scene. Subjects press the left (right) mouse button if the building shown in the top center is towards the left (right) of the test scene.

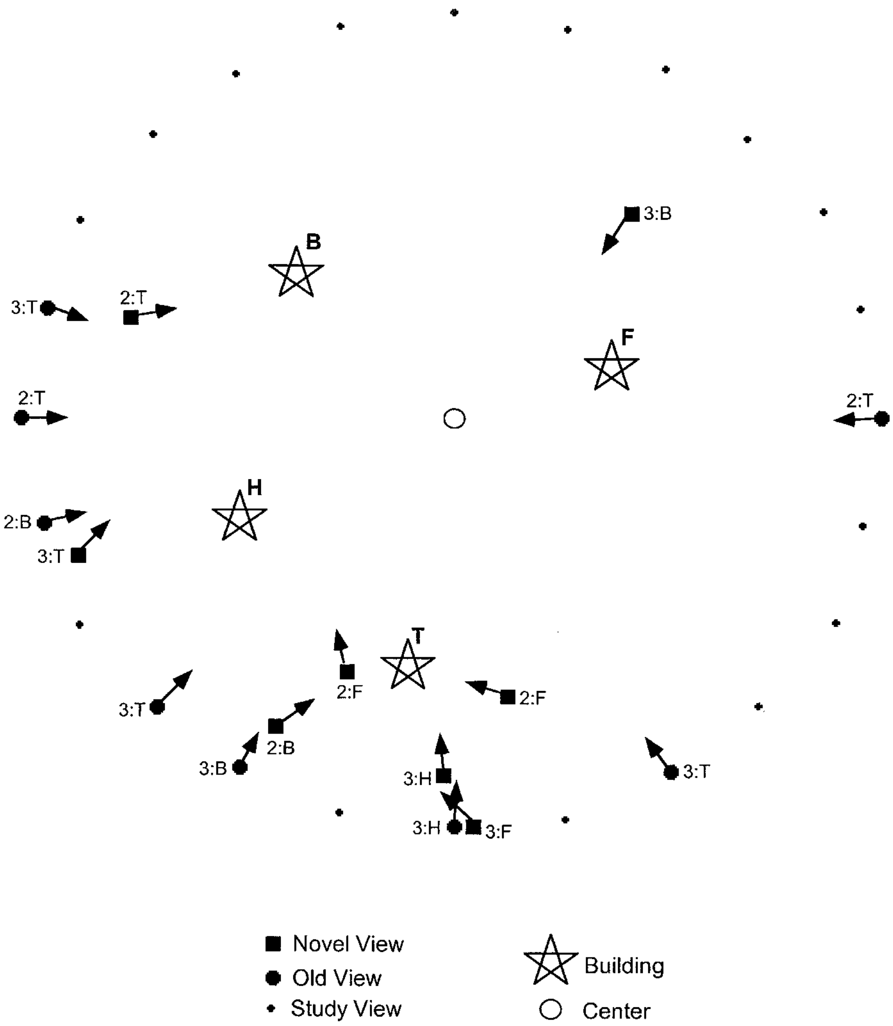

Figure 3. Vantage points for the novel and old test scenes for environment No. 2. Also indicated are the number of buildings visible in each scene (2 or 3) and the target building (B, F, H or T).

Experiment 1: Results

To examine the effects of exposure time, presentation order, and a frame of reference on learning relative directions between landmarks, a 2 (self pace vs. fixed pace) $\times$ 2 (sequential vs. random order) $\times$ 2 (trees vs. no trees environment) mixed ANOVA was performed, with repeated measures on the last two factors. Viewpoint dependent and viewpoint independent representations were examined by analyzing accuracy and RT (response time) for old and novel test scenes.

Overall, 86.9% of all direction responses were correct, well above 50% chance performance, with an average RT of 4.3s. There were no sex-related differences for RT ($F(1,30) = 1.17$, $p > 0.10$), but males were significantly more accurate than females, 93% vs. 82% respectively ($F(1,30) = 7.51$, $p < 0.05$).

Self-pace vs. fixed-pace learning

Mean exposure time (ET) per scene during the self-pace learning condition did not vary across the four different spatial environments, suggesting that they were of comparable difficulty. However, ET was significantly greater during the random presentation order (6.5 s) than the sequential presentation order (4.8 s), $F = 59.8$, $p = 0.0001$. During debriefing, subjects mentioned that they needed extra time to re-orient each view to some dominant orientation. There was no significant difference in ET for the frame of reference conditions (tree and no-tree environments); nor interactions with the presentation order. ET was generally longer for the first few snapshots, a trend that was somewhat stronger during the random presentation order than the sequential presentation order.

Overall, there was no between-subjects main effect for learning condition (self-pace vs. fixed-pace) on RT ($F(1,30) = 1.33$, $p > 0.10$), but there was a trend towards greater accuracy in the self-pace learning condition (91.0% vs. 82.8%; $F(1,30) = 3.52$, $0.05 < p < 0.10$). Even though subjects in the self-pace learning condition had spent considerable more time to learn the spatial configuration of the buildings, they were not significantly faster in making their direction estimations and they were only marginally more accurate.

The interaction between old and novel scenes with learning condition was not significant for accuracy ($F(1,30) = 0.03$, $p > 0.10$), nor for RT ($F(1,30) = 0.58$, $p > 0.10$). Subjects in both learning conditions were significantly faster in responding to old views than to novel views.

Random vs. sequential presentation order

There was no within-subjects main effect of presentation order (random vs. sequential) on RT ($F(1,30) = 1.53$, $p > 0.10$) nor on accuracy ($F(1,30) = 0.03$, $p > 0.10$). Subjects were able to make relative direction estimations with the same accuracy and speed regardless of the manner in which the scenes were presented to them during learning. There was also no significant interaction between presentation order (random vs. sequential) and learning condition (self pace vs. fixed pace) for RT ($F(1,30) = 0.20$, $p > 0.10$) nor for accuracy

($F(1,30) = 1.38$, $p > 0.10$). These findings suggest that subjects learned the relative directions between landmarks equally well regardless of the manner in which they viewed the environment (random or sequential) or how much time was spent learning the environment (self-pace or fixed-pace).

There was no interaction between test scenes (old vs. novel) and presentation order (random vs. sequential) for accuracy ($F(1,30) = 0.07$, $p > 0.10$) nor for RT ($F(1,30) = 0.11$, $p > 0.10$). Therefore, the presentation of scenes did not have an effect on whether a viewpoint dependent or viewpoint independent representation was formed.

Frame of reference

Overall, there was no within-subject main effect for the trees vs. no trees environments on accuracy ($F(1,30) = 2.13$, $p > 0.10$) nor on RT ($F(1,30) = 2.62$, $p > 0.10$). However, when comparing responses to test scenes from the trees environments in which the trees were actually visible (66% of the test scenes) with responses from the non-tree learning condition, there was a main effect on RT ($F(1,31) = 6.00$, $p < 0.05$) but not on accuracy ($F(1,31) = 1.56$, $p > 0.10$). Essentially, when the trees were visible during testing, subjects were significantly faster (but not more accurate) in making direction judgments.

The interaction between presentation order (sequential vs. random) and the frame of reference condition (trees vs. no trees) was not significant for RT ($F(1,30) = 0.21$, $p > 0.10$) nor for accuracy ($F(1,30) = 0.49$, $p > 0.10$). Subjects were able to learn the relative directions between landmarks equally with and without a frame of reference, and irrespective of the manner in which the scenes were presented. This same pattern was also observed when comparing only those test scenes in which the trees were visible.

There was also no interaction between test scenes (old vs. novel) and frame of reference condition (trees vs. no trees) for RT ($F(1,30) = 0.02$, $p > 0.1$), but there was a significant interaction for accuracy ($F(1,30) = 5.20$, $p < 0.05$). The same pattern was found when analyzing accuracy only for test scenes in which the trees were visible ($F(1,30) = 10.24$, $p < 0.001$; see Figure 4). When subjects had access to a frame of reference they were more accurate in their direction estimations only on old test scenes, not on novel test scenes. This finding suggests that subjects may use different types of information when solving direction estimations in old and novel scenes.

Old vs. novel views

A repeated measures ANOVA showed a significant within-subjects main effect of test scene (old vs. novel) on RT ($F(1,30) = 49.16$, $p < 0.001$), but not on accuracy ($F(1,30) = 0.92$, $p > 0.10$). Mean RT and accuracy for

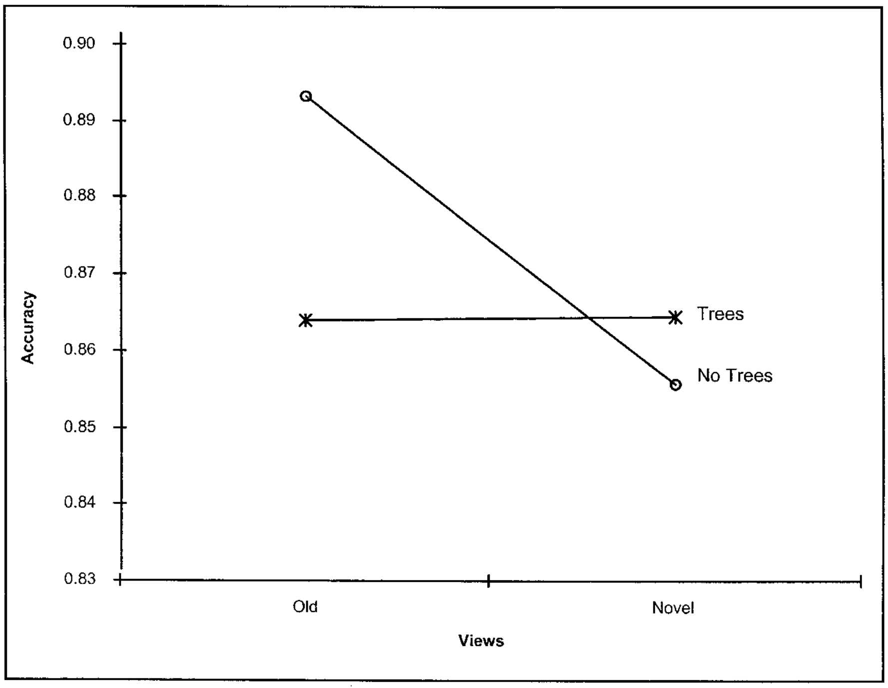

Figure 4. Significant interaction for accuracy between trees/no trees learning condition and old/novel test scenes (using old and novel test scenes with trees visible).

old views were 4,064 ms and 87.4%, and 4,500 ms and 86.4% for novel views. Before we can conclude that old views have faster responses, we need to control for the number of buildings visible in each scene as it is possible that scene contents affects response time. Indeed, subjects were significantly faster (3,974 vs. 4,643 ms RT) and more accurate (88.5% vs. 85.2% correct) when responding to test scenes with 3 buildings than with 2 buildings ($F(1,31) = 36.17$, $p < 0.001$, $F(1,31) = 5.81$, $p < 0.05$, respectively). This pattern was also present for old and novel test scenes separately. However, novel test scenes still required significantly more time, even after controlling for the number of buildings ($F(1,31) = 20.83$, $p < 0.001$ for 2-building test scenes; $F(1,31) = 4.03$, $p = 0.05$ for 3-building test scenes; see Figure 5). There was no such effect of old vs. novel views on accuracy.

Experiment 2: Single study view

The purpose of Experiment 2 is to determine if the difference in RT between old and novel views in Experiment 1 was due to anything other than their

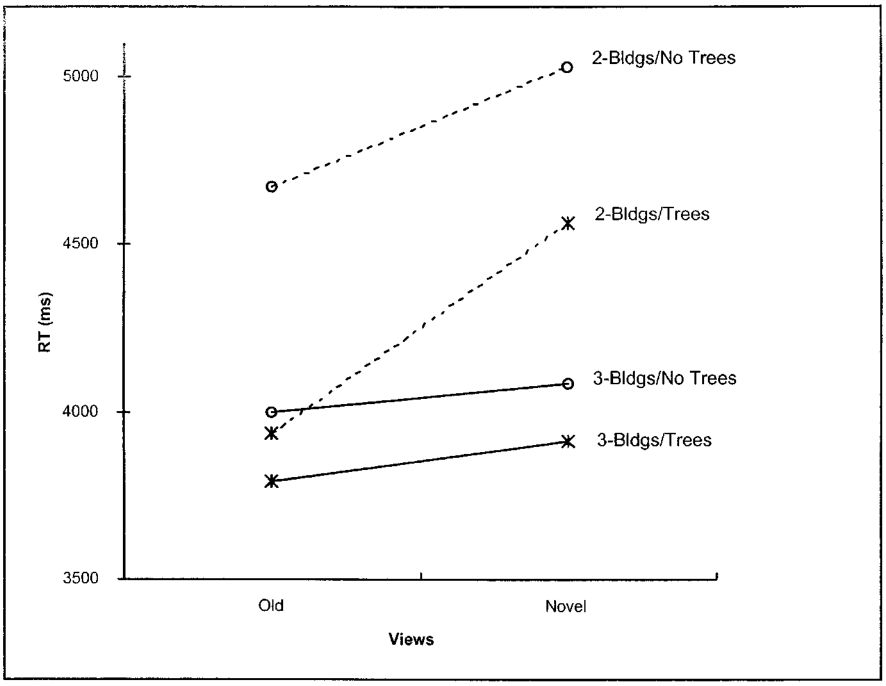

Figure 5. Reaction times for novel and old test scenes shown separately for scenes with 2 or 3 buildings visible, and with or without trees visible.

familiarity. Therefore, Experiment 2 involves studying each environment from a single vantage point and comparing direction judgments on the old and novel views.

Subjects

A total of 16 subjects (12 male and 4 female) participated in Experiment 2. Subjects were recruited from the Cambridge Basic Research subject pool and were compensated for their participation. Subjects were run in individual sessions lasting 30 minutes. None of the subjects who participated in Experiment 2 also participated in Experiment 1.

Stimuli and procedures

The same stimuli used in Experiment 1 were also used in Experiment 2. Subjects in Experiment 2 only saw a single view showing all four buildings. After studying this view for 3 minutes, subjects were shown 16 test scenes (note that the distinction between old and novel test views is not meaningful here, as all test views are novel). Subjects were tested only on the four environments that did not have a line of trees. Response times that were more than 2 standard deviations above the mean (15.4s) were excluded from the analysis. This resulted in the exclusion of 43 responses out of a total of 1,024 (4.20%).

Experiment 2: Results

Mean RT was 5.57s, which is considerably slower than the RT of 4.67s in the comparable condition from Experiment 1 (fixed-pace – no tree condition), although its statistical significance is only marginal ($F(1,30) = 3.38$, $p = 0.076$). Mean accuracy was 87.6%, which is slightly higher than the 81.8% accuracy in Experiment 1 (fixed-pace – no tree condition), but not significantly so. No significant differences in accuracy or RT were observed between old and new views (defined in Experiment 1), $t(16) = 0.01$, $p = 0.99$, $t(16) = 0.42$, respectively.

Although the test views in this experiment were of course all novel, half of them were old views from Experiment 1, while the other half were novel views from Experiment 1. This allowed us to determine whether old and novel test views perhaps differed in some way that could have caused the difference in RT between old and novel scenes observed in Experiment 1. Since neither accuracy nor RT differed for the old and novel test scenes in the single-study view experiment, it follows that any differences found in the multiple-study experiments can be attributed to familiarity or novelty of the views.

Conclusions and discussion

The main results of this study can be summarized as follows. First, learning relative directions between landmarks in a simple virtual environment consisting of four buildings was rather robust. Whether a single view was studied for a long time or multiple views from around the environment were each studied briefly in an sequential or a random order and under a self-paced or fixed-paced condition, final direction judgments were always about 85% correct (with 50% being chance level). Only under the self-paced condition was accuracy slightly and marginally higher. This finding suggests that it is possible to accurately learn relative directions between a small number of landmarks in a desktop virtual environment in a fairly short amount of time, under varying viewing conditions.

Second, the order in which environmental scenes were viewed did not affect the ability to encode spatial relations between landmarks. Subjects were equally accurate in their relative direction estimations in both the random and sequential presentation orders. This result may suggest that subjects do not organize views of the environment in a sequential manner. Rather, it appears as though subjects store a loose collection of distinct images of the environment in memory. Therefore, the manner in which an environment is perceived may have little, if any impact on the acquisition of spatial relations among landmarks in a desktop virtual environment.

Third, viewing a configuration of landmarks in a desktop virtual environment from multiple perspectives produces a viewpoint dependent representation in memory. Subjects had significantly greater response times for new viewing perspectives, as compared to previously viewed scenes. However, there was no difference in the accuracy of the relative direction estimations for old and new scenes. Therefore, even though subjects appear to mentally represent the environment as a collection of distinct images, it only affects their response time for relative direction estimations, not accuracy. Not only is a viewpoint dependent representation acquired from studying inter-object spatial relations from two (Shelton and McNamara 1997) or four perspectives (Diwadkar and McNamara 1997), but it is also acquired from studying a single environment from 24 different perspectives.

Fourth, a viewpoint dependent representation of the environment persists despite learning under conditions of spatio-temporal discontinuity and changes to a frame of reference. This was seen in a lack of an interaction for response time between the presentation order and tree environments and between presentation order and test scenes. The dominance of viewpoint dependence is so strong that subjects adopt a viewer-centered frame of reference even when they were given an external frame of reference. The manner in which the environment is viewed or the contents of the scene do not impact the formation of a viewpoint dependent representation. However it should be noted that the frame of reference does help somewhat in accuracy of relative direction estimations for old scenes only.

In the future, we hope to extend our research by examining how individuals learn to navigate through immersive virtual environments. Specifically, we hope to better understand the role that vestibular and proprioceptive information plays in the formation of a viewpoint dependent or independent representation. Ultimately, this research will lead towards a more complete picture of human spatial cognition in both real and virtual environments.

References

Albert, W.S., Reinitz, M.T., Beusmans, J.B. and Gopal, S. (1999). The Role of Attention in Spatial Learning During Simulated Navigation, *Environment and Planning A* 31: 1459–1472.

Allen, G.L. (1988). The Acquisition of Spatial Knowledge under Conditions of Temporospatial Discontinuity, *Psychological Research* 50: 183–190.

Allen, G.L., Siegel, A.W. and Rosinski, R.R. (1978). The Role of Perceptual Context in Structuring Spatial Knowledge, *Journal of Experimental Psychology: Human, Learning, and Memory* 4: 617–630.

Couclelis, H., Golledge, R.G., Gale, N. and Tobler, W. (1987). Exploring the Anchor-Point Theory of Spatial Cognition, *Journal of Environmental Psychology* 7: 99–122.

Darken, R.P., Allard, T. and Achille, L.B. (1998). Spatial Orientation and Wayfinding in Large-Scale Virtual Spaces: An Introduction, *Presence* 7: 101–107.

Diwadkar, V.A. and McNamara, T.P. (1997). Viewpoint Dependence in Scene Recognition, *Psychological Science* 8: 302–307.

Gärling, T. and Golledge, R.G. (eds.) (1993). *Behavior and Environment: Psychological and Geographical Approaches*. Amsterdam: Elsevier Science Publishers.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.

McNamara, T.P. and Diwadkar, V.A. (1996). Viewpoint Dependence in Human Spatial Memory, *Cognitive and Computational Models of Spatial Representation: Papers from the 1996 AAAI Spring Symposium*. AAAI Press: Menlo Park, CA.

Montello, D.R. (1992). Scale and Multiple Psychologies of Space. In A.U. Frank, I. Campari and U. Formentini (eds.), *International Conference: GIS – From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning*. Pisa, Italy.

Richardson, A.R., Montello, D.R. and Hegarty, M. (1999). Spatial Knowledge Acquisition from Maps, and from Navigation in Real and Virtual Environments, *Memory and Cognition* 27: 741–750.

Shelton, A.L. and McNamara, T.P. (1997). Multiple Views of Spatial Memory, *Psychonomic Bulletin and Review* 4: 102–106.

Subbiah, I., Veltri, L., Liu, A. and Pentland, A. (1996). Paths, Landmarks, and Edges as Reference Frames in Mental Maps of Simulated Environments. *Cambridge Basic Research Technical Report, 96-11*, Cambridge, MA.

Tlauka, M. and Wilson, P.N. (1996). Orientation-Free Representations from Navigation Through a Computer-Simulated Environment. *Environment and Behavior* 28: 647–664.