Psychonomic Bulletin & Review
2008, 15 (2), 322-327
doi: 10.3758/PBR.15.2.322

Spatial memories of virtual environments: How egocentric experience, intrinsic structure, and extrinsic structure interact

Jonathan W. Kelly and Timothy P. McNamara
Vanderbilt University, Nashville, Tennessee

Previous research has uncovered three primary cues that influence spatial memory organization: egocentric experience, intrinsic structure (object defined), and extrinsic structure (environment defined). In the present experiments, we assessed the relative importance of these cues when all three were available during learning. Participants learned layouts from two perspectives in immersive virtual reality. In Experiment 1, axes defined by intrinsic and extrinsic structures were in conflict, and learning occurred from two perspectives, each aligned with either the intrinsic or the extrinsic structure. Spatial memories were organized around a reference direction selected from the first perspective, regardless of its alignment with intrinsic or extrinsic structures. In Experiment 2, axes defined by intrinsic and extrinsic structures were congruent, and spatial memories were organized around reference axes defined by those congruent structures, rather than by the initially experienced view. The findings are discussed in the context of spatial memory theory as it relates to real and virtual environments.

Memories for learned environments play a critical role in many spatial tasks, from selecting an alternative route home during rush hour to finding the kitchen in the dark. Much of the relevant experimental work has focused on spatial memory organization, and a nearly ubiquitous finding in this research has been that spatial memories are represented with respect to one or two reference directions (see McNamara, 2003, for an overview). These reference directions are selected through a combination of egocentric experience within the learned environment and properties of the environmental structure (e.g., Shelton & McNamara, 2001). Both can be considered cues that influence the reference system used to represent spatial memories. Three broad categories of these cues have emerged from this research: egocentric experience, extrinsic structure, and intrinsic structure.

Because the body is typically used to act on the environment that is to be remembered, selecting the primary reference axis from the learning perspective seems prima facie to be a natural way to organize spatial memories. In fact, early results indicated that egocentric experience is dominant under certain conditions. After object locations have been learned from multiple perspectives, subsequent recall and recognition is best for the learned perspectives and degrades with increasing angular distance from those learned perspectives (Diwadkar & McNamara, 1997; Shelton & McNamara, 1997). Whereas some experiments have demonstrated that spatial memories are organized around multiple reference axes based on multiple experienced views, others have shown that a single reference direction is established from the initial learning perspective and that this organization persists even after extensive experience with other perspectives (Avraamides & Kelly, 2005; Kelly, Avraamides, & Loomis, 2007).

Other work has demonstrated an important role for extrinsic structure, whereby salient environmental features influence reference frame selection. Werner and Schmidt (1999; see also Montello, 1991) found that residents of a German city were best at recalling relative object locations from imagined perspectives parallel to major city streets, as compared with oblique perspectives. Other environmental features, such as lakes and nearby buildings, can have a similar impact on spatial memories (McNamara, Rump, & Werner, 2003). Furthermore, extrinsic structure has been shown to interact with egocentric experience in predictable ways. Shelton and McNamara (2001) had participants learn object locations from two perspectives within a rectangular room. One perspective was aligned and one was misaligned with the extrinsic axes redundantly provided by the room walls and a square mat on the floor. The sequence of experienced perspectives was varied so that half of the participants learned the misaligned view first and the aligned view second and the other half received the reverse viewing order. Subsequent retrieval of relative object locations was best when the aligned learning perspective was imagined. Performance when the misaligned learning perspective was imagined was no better than that for novel perspectives that had never been experienced, and this held true regardless of the order in which the aligned and the misaligned perspectives were

J. W. Kelly, jonathan.kelly@vanderbilt.edu

experienced during learning. However, when the extrinsic structures defined by the room and mat were misaligned with one another by 45º, performance was best when the initially experienced view was imagined, regardless of the learning perspective order. Thus, egocentric experience appears to dominate in the absence of a single coherent extrinsic structure. When a reliable extrinsic axis is present and learning occurs from a perspective aligned with that axis, spatial memories are influenced by the extrinsic axis, rather than by egocentric experience. Interestingly, our daily environments commonly have multiple potential extrinsic structures misaligned with one another, such as streets, tree lines, waterfronts, and buildings. If these separate environmental structures compete with each other, egocentric cues may predominate.

In addition to environmental features, the layout of the objects themselves can provide a salient intrinsic structure. Mou and McNamara (2002) had participants learn a regular arrangement of objects organized in rows and columns, viewed from a single learning perspective oblique to the intrinsic axes. The object layout consisted of two primary orthogonal axes defined by the object locations. During learning, the experimenter emphasized the intrinsic structure of the layout through verbal instruction, and subsequent testing showed that the objects were represented along reference axes defined by the intrinsic rows and columns. This occurred despite the fact that the observers never actually experienced a perspective aligned with the intrinsic axes.

The aim of the experiments reported here was to further the understanding of how cues provided by egocentric experience, extrinsic structure, and intrinsic structure interact when a spatial layout is learned. Although previous experiments have highlighted the individual importance of each cue, it is unknown how spatial memories are formed when all three are available during learning. This represents an important step toward understanding the organization of spatial memories for our daily environments, which commonly contain numerous potential reference frames. In these experiments, spatial layouts were learned in immersive virtual reality (VR), in which movement through the virtual environment was achieved by physically turning and walking, as in a real environment. VR has previously been used to replicate and extend real-world findings on spatial memory retrieval (e.g., Kelly et al., 2007; Williams, Narasimham, Westerman, Rieser, & Bodenheimer, 2007). However, the use of VR in studying spatial memory encoding has produced mixed results. Some studies have reported substantial differences between spatial memories acquired through real-world navigation and those learned through desktop VR, in which movement through the virtual environment was controlled by a joystick or a keyboard (Richardson, Montello, & Hegarty, 1999). But these findings may have been due to the unnatural movement through desktop VR, as compared with the physical rotations and translations that accompany real-world exploration. Indeed, other work has indicated that complex foraging tasks requiring accurate spatial memories can benefit greatly from physical movement and that physically walking while navigating through a virtual environment (Ruddle & Lessels, 2006) produces performance on par with performance in real environments (Lessels & Ruddle, 2005). Further work using immersive VR has indicated the importance of egocentric experience when objects are learned in a virtual environment devoid of strong environmental axes (Kelly et al., 2007). This finding suggests that spatial memories of virtual environments may be structured similarly to memories for real environments.

EXPERIMENT 1

Experiment 1 was designed to examine how egocentric experience, extrinsic structure, and intrinsic structure affect reference frame selection when all three cues are available during learning. Participants learned a layout of objects in a virtual environment that contained two sets of environment-defined axes: one defined by the columnar organization of the virtual objects (intrinsic), and the other defined by the square walls of the virtual room (extrinsic). These two sets of axes were misaligned with one another by 45º, and this environment will be referred to as the *incongruent* environment (Figure 1, left panel). All the participants studied the objects from two perspectives, one that was aligned with the intrinsic structure but misaligned with the extrinsic structure (the 0º view in Figure 1, left) and one that was aligned with the extrinsic structure but misaligned with the intrinsic structure (the 135º view in Figure 1, left), and viewing order was manipulated. If either the intrinsic or the extrinsic structure was more salient than the other, recall should be best for perspectives aligned with that salient reference frame, regardless of the viewing order. If, on the other hand, the conflicting intrinsic and extrinsic structures negated the influence of one another, egocentric experience should dominate, and the preferred reference axis should correspond to the initially experienced view.

In addition, Experiment 1 was designed to provide supporting evidence for current spatial memory theories, using virtual environments. In order for VR to be a useful tool for studying complex spatial behaviors such as navigation and route planning, spatial memories for simple, well-controlled virtual environments should be organized similarly to spatial memories for real environments. To

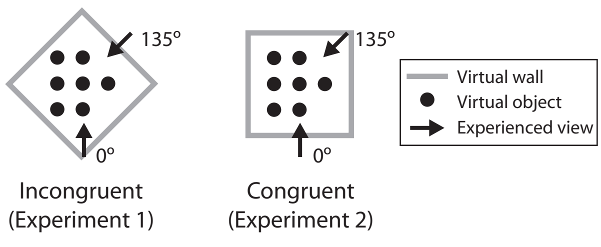

Figure 1. Stimuli used in Experiment 1 (left panel) and Experiment 2 (right panel).

our knowledge, this is the first thorough investigation of reference frames in spatial memories for immersive virtual environments.

Method

Participants

Twenty-four participants (12 of them male) from the Nashville community participated in exchange for monetary compensation.

Stimuli and Design

The stimuli were presented through an nVisor SX (from NVIS, Reston, VA) head-mounted display (HMD). The HMD presented stereoscopic images at 1,280 × 1,024 resolution, refreshed at 60 Hz. The HMD field of view was 47º horizontal × 38º vertical. Graphics were rendered by a 3.0-GHz Pentium 4 processor with a GeForce 6800 GS graphics card using Vizard software (WorldViz, Santa Barbara, CA). Head orientation was tracked using a three-axis orientation sensor (InertiaCube2 from Intersense, Bedford, MA), and head position was tracked using a passive optical-tracking system (Precision Position Tracker, PPT X4 from WorldViz, Santa Barbara, CA). Graphics displayed in the HMD were updated on the basis of sensed position and orientation of the participant's head.

The virtual environment consisted of seven virtual objects (shark, ball, train, bug, cup, lamp, and plant) placed on identical green pillars that were 60 cm tall. Each object was scaled to fit within a 20-cm cube. The objects were arranged along a rectilinear grid, evenly spaced every 75 cm (see Figure 1). The objects were arranged in three columns, the left and middle column containing three objects and the right column containing one object. In addition, a square (3 m) virtual room surrounded the scene. The room was rotated about its vertical axis to be misaligned by 45º with the intrinsic structure. Room walls were covered with a brick pattern containing strong texture gradient depth cues, and the floor was covered with a random carpet texture.

All the participants learned the object locations from two perspectives, 0º and 135º (corresponding to two unique viewing locations; see Figure 1), and viewing order was manipulated. Figure 2 shows the participants' perspective of the incongruent environment from 135º. After learning, the participants were removed from the virtual environment and performed judgments of relative direction (JRDs) in which they were asked to imagine standing at the location of one object, facing a second object, and then point to a third object from that imagined perspective. JRD trials were presented on a 19-in. computer monitor, and pointing responses were made using a joystick (Freedom 2.4 by Logitech, Fremont, CA) placed on the desk in front of the participants.

The independent variables were viewing order (0º and then 135º or 135º and then 0º) and imagined perspective in the JRD task. Viewing order was manipulated between participants, and imagined perspective was manipulated within participants. JRD stimuli were selected to test eight different imagined perspectives, spaced every 45º from 0º to 315º. The objects appeared as standing, orienting, and pointing objects with equal frequency. For each imagined perspective, 6 trials were constructed to create correct egocentric pointing directions of 45º, 90º, 135º, 225º, 270º, and 315º, resulting in a total of 48 trials.

The dependent measures were pointing error (the absolute angular error of the pointing response) and pointing latency (the latency between presentation of the three object names for a given trial and completion of the pointing response).

Procedure

Learning phase. The experimenter met the participant in front of the learning room. After providing informed consent, the participant donned the HMD while standing in the hallway. Once the HMD was in place, the participant was led into the lab and positioned for the first view (either 0º or 135º). During transport, the blank HMD display served as a blindfold, blocking any view of the physical lab structure. In addition, the lab lights were turned off to prevent the participant's viewing the lab around the edges of the HMD.

Once the participant was properly positioned at the first viewing location and the HMD was turned on, the experimenter named each of the test objects in a random sequence. To direct the participant's attention to a specific location, the experimenter pressed a key to temporarily change the color of the pillar underneath the object from green to white. After all of the objects had been named, the participant was instructed to study the layout for 30 sec. During learning,

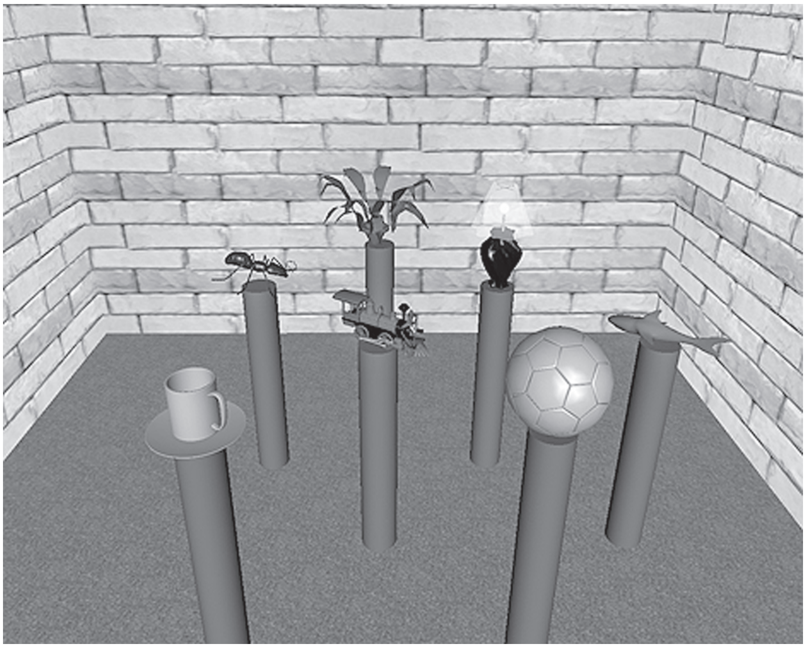

Figure 2. The participant's view of the incongruent environment from the 135º perspective. Some objects were enlarged for publication purposes.

the participants were told not to move from the study location but that they were free to turn their heads to look around, and the participants regularly did so in order to study all objects. Following this study interval, the objects were hidden, but the pillars on which they rested remained. The experimenter then indicated each pillar in a random sequence and asked the participant to name the corresponding object from that location. This learning sequence was repeated until the participant successfully named all the objects twice. Typically, this occurred after two or three learning intervals.

After learning from the first location was completed, the display was turned off (serving as a blindfold during transport), and the participant walked, guided by the experimenter, to the second viewing location (approximately 3 m from the first viewing location). When the display was turned back on, the participant's view of the virtual environment reflected the physical movements he or she had undergone during transport between the first and the second viewing locations. This procedure simulated that in Shelton and McNamara (2001), in which participants walked blindfolded from the first to the second viewing location. After completion of the same learning procedure from the second perspective, the display was turned off again, the participant was led back out into the hallway, and the HMD was subsequently removed.

Testing phase. After learning, the participant was led to another room on the same floor of the building to complete the JRD task. The participant was seated in front of a computer monitor and completed 4 practice JRD trials in which either buildings on the Vanderbilt campus or U.S. cities were used, depending on participant familiarity. After practice, the participant completed 48 trials composed of previously studied objects. Each trial was initiated when the participant pressed a button on the joystick, at which point the imagined standing object, facing object, and target object were simultaneously displayed in text (e.g., "Imagine standing at the train, facing the shark. Point to the plant."). Trials were pseudorandomized, with the constraint that the same perspective was never tested twice in a row. The participant responded by moving the joystick in the direction of the target object from the imagined perspective. Each response was recorded when the joystick was deflected by 30º from vertical.

Results

Pointing error was more responsive to the independent variables than was pointing latency, which was largely unaffected.[1] In the interest of brevity, we will focus on the error data. Pointing error (plotted in Figure 3 as a function of imagined perspective) was analyzed in a 2 (gender) × 2 (viewing order) × 8 (imagined perspective) mixed-model ANOVA. The significant main effect of imagined perspective [$F(7,140) = 2.77, p = .01, \eta_p^2 = .12$] was qualified by a significant interaction between imagined perspective and viewing order [$F(7,140) = 2.18, p = .04, \eta_p^2 = .10$]. No other main effects or interactions were significant. The participants were more accurate for the imagined perspective corresponding to the first learning perspective (0º for the 0º–135º learning order and 135º for the 135º–0º learning order) than for that corresponding to the second learning perspective, as indicated by the interaction contrast [$F(1,20) = 6.03, p = .023, \eta_p^2 = .23$]. There was no overall benefit for imagined perspectives aligned with either the intrinsic (i.e., the 0º, 90º, 180º, and 270º perspectives) or the extrinsic (i.e., the 45º, 135º, 225º, and 315º perspectives) axes.

Discussion

In Experiment 1, the incongruent intrinsic and extrinsic environmental structures resulted in spatial memories with

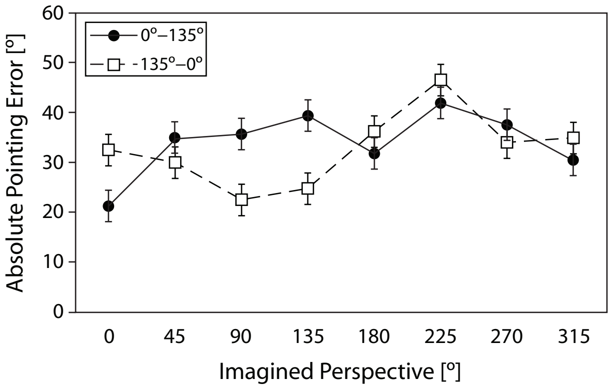

Figure 3. Absolute pointing error in Experiment 1 after the incongruent environment was learned, plotted separately for 0º–135º and 135º–0º viewing orders. Error bars are standard errors estimated from the ANOVA.

a preferred reference direction selected from the initially experienced perspective. Although one learning perspective was aligned with the intrinsic structure and one was aligned with the extrinsic structure, neither had a consistent impact across the two viewing orders. These results parallel findings by Shelton and McNamara (2001), who manipulated learning perspective order in the presence of two incongruent extrinsic structures defined by the room walls and a mat on the floor. Under these conditions, they found that the primary reference direction used to structure spatial memories was selected from the initially viewed perspective, regardless of the viewing order. This provides initial evidence that intrinsic and extrinsic structures exert similar influences on spatial memory and that neither is preferentially used when they are placed in competition.

By replicating and extending a primary finding on reference frame congruency, Experiment 1 also serves to further validate the use of immersive VR technology for studying spatial memory. This finding provides further support for the use of VR, which has been previously used to demonstrate the role of egocentric experience (Kelly et al., 2007), but not environmental structure.

EXPERIMENT 2

Experiment 2 was designed to investigate the roles of egocentric experience, intrinsic structure, and extrinsic structure when the latter two were congruent with one another. To achieve this, the axes defined by the object layout and the room walls were placed in alignment, and this environment will be referred to as the *congruent* environment (see Figure 1, right). The results from Experiment 1 indicated that intrinsic and extrinsic structures interact with each other in a manner functionally equivalent to two extrinsic structures. As such, we anticipated that congruent intrinsic and extrinsic structures would produce spatial memories organized around their congruent axes, regardless of viewing order. This prediction was based on previous results when environments with congruent extrinsic structures were used (Shelton & McNamara, 2001).

Method

Participants

Twenty-four participants (12 of them male) from the Nashville community participated in exchange for monetary compensation. One female participant was replaced due to failure to follow experimenter instructions.

Stimuli, Design, and Procedure

The design and procedure were identical to those in Experiment 1. The stimuli were modified so that intrinsic and extrinsic structures defined congruent axes (see Figure 1, right). The 0º learning view was aligned with both structures, and the 135º view was misaligned with both structures. Viewing order was manipulated between participants, and imagined perspective was manipulated within participants.

Results

Pointing error (plotted in Figure 4) was analyzed in a 2 (gender) × 2 (viewing order) × 8 (imagined perspective) mixed-model ANOVA. This analysis revealed only a main effect of imagined perspective [$F(7,140) = 9.71$, $p < .001$, $\eta_p^2 = .33$], where performance was best on perspectives aligned with the intrinsic and extrinsic structures (0º, 90º, 180º, and 270º), as compared with misaligned perspectives (45º, 135º, 225º, and 315º) [$F(1,20) = 53.17$, $p < .001$, $\eta_p^2 = .73$]. No other main effects or interactions were significant.

Discussion

When axes defined by intrinsic and extrinsic structures were congruent with one another, spatial memories were organized around the orthogonal reference axes redundantly defined by the environment. Performance for the 135º perspective was no better than that for other novel misaligned perspectives, and performance for the 0º perspective was no better than that for other novel aligned perspectives, indicating that memories were organized around environmentally defined axes and were not based strictly on egocentric experience. This replicates previous findings with congruent extrinsic structures (Shelton & McNamara, 2001) and serves to further validate immersive VR as a tool for studying spatial memory.

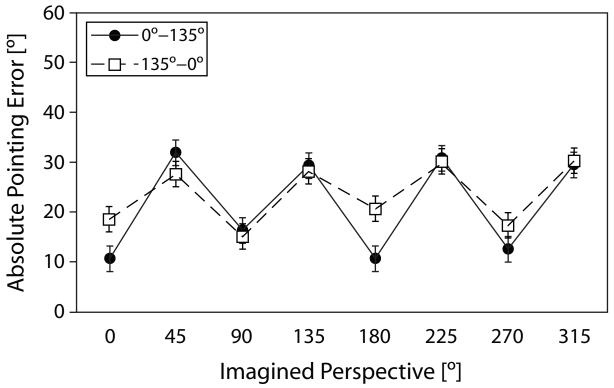

Figure 4. Absolute pointing error in Experiment 2 after the congruent environment was learned, plotted separately for 0º–135º and 135º–0º viewing orders. Error bars are standard errors estimated from the ANOVA.

GENERAL DISCUSSION

The primary purpose of these experiments was to evaluate the interaction between egocentric experience, intrinsic structure, and extrinsic structure in a spatial memory task. Previous work has indicated that the effect of multiple extrinsic structures (such as the room walls and a mat on the floor) depends on their congruency with one another. Shelton and McNamara (2001) found that when extrinsic structures were incongruent, spatial memories were organized around a reference direction selected from the initially experienced perspective. When the same extrinsic structures were congruent, spatial memories were organized around the redundantly defined axes, rather than the initially experienced view. Likewise, incongruent intrinsic and extrinsic structures in Experiment 1 resulted in egocentric selection of a reference direction from the initially learned perspective. Neither intrinsic nor extrinsic structure was preferentially used when the incongruent environment was learned. Instead, incongruency between multiple types of environmental structures resulted in egocentric spatial memories. Because environments often contain multiple potential reference frames, such as roads, trees, and hills, it is possible that egocentric experience predominates in spatial memories of our daily environments. However, further research is needed to determine how differences in spatial memory organization might lead to differences in navigation performance within those environments. In addition, the intrinsic structure provided by the columnar object organization in these experiments is rather artificial, and it remains to be seen how more naturalistic structures would influence spatial memories.

In Experiment 2, when intrinsic and extrinsic structures were congruent with one another, spatial memories were organized around the congruent environmental axes, rather than egocentric experience. We hypothesize that when learning first occurred from a perspective misaligned with the congruent intrinsic and extrinsic structures (the 135º learning view in Experiment 2), the participants selected an egocentric reference direction from that misaligned perspective. However, subsequent learning from an aligned perspective (the 0º learning view in Experiment 2) resulted in restructuring of the memory around the redundantly provided intrinsic and extrinsic axes. Recent experiments have shown that this restructuring occurs within 10–30 sec of learning from the aligned view (Valiquette, McNamara, & Lebrecque, 2007).

Previous experiments using immersive VR have shown that spatial memories for virtual environments lacking strong extrinsic or intrinsic structure are heavily influenced by egocentric experience (Kelly et al., 2007), an important replication of similar findings in real environments (Diwadkar & McNamara, 1997; Shelton & McNamara, 1997). The present experiments extend this work to more complex virtual environments, containing multiple environmental structures. On the basis of the present evidence, it appears that spatial memories for virtual environments are organized similarly to memories for real environments and are similarly responsive to environment structure and egocentric experience. VR holds obvious advantages for studying spatial memory, due to the ease with which novel environments can be created, from room-sized to city-sized environments. The present experiments, however, addressed only room-sized environments. Other studies in which joystick-based locomotion has been used have shown substantial differences in spatial memories for real and virtual large-scale environments (Richardson et al., 1999). It is unclear whether these divergent findings result from differences in environmental scale or differences in exploration mode, and further work is needed to disentangle these factors. Furthermore, learning in these experiments occurred only from two perspectives, whereas natural exploration typically involves learning from many perspectives. It remains to be seen how these results will extend to more complex learning conditions.

AUTHOR NOTE

This work was supported by NIMH Grant 2-R01-MH57868 and a Vanderbilt University Discovery Grant. We thank Nora Newcombe and two anonymous reviewers for helpful feedback on a previous draft. Correspondence concerning this article should be addressed to J. W. Kelly, Department of Psychology, Vanderbilt University, 111 21st Avenue South, Nashville, TN 37203 (e-mail: jonathan.kelly@vanderbilt.edu).

REFERENCES

AVRAAMIDES, M. N., & KELLY, J. W. (2005). Imagined perspective-changing within and across novel environments. In C. Freksa, B. Nebel, M. Knauff, & B. Krieg-Brückner (Eds.), *Lecture notes in artificial intelligence: Spatial cognition IV* (pp. 245-258). Berlin: Springer.

DIWADKAR, V. A., & MCNAMARA, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, **8**, 302-307.

KELLY, J. W., AVRAAMIDES, M. N., & LOOMIS, J. M. (2007). Sensorimotor alignment effects in the learning environment and in novel environments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 1092-1107.

LESSELS, S., & RUDDLE, R. A. (2005). Movement around real and virtual cluttered environments. *Presence: Teleoperators & Virtual Environments*, **14**, 580-596.

MCNAMARA, T. P. (2003). How are the locations of objects in the environment represented in memory? In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Lecture notes in artificial intelligence: Spatial cognition III* (pp. 174-191). Berlin: Springer.

MCNAMARA, T. P., RUMP, B., & WERNER, S. (2003). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonomic Bulletin & Review*, **10**, 589-595.

MONTELLO, D. R. (1991). Spatial orientation and the angularity of urban routes: A field study. *Environment & Behavior*, **23**, 47-69.

MOU, W., & MCNAMARA, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 162-170.

RICHARDSON, A. E., MONTELLO, D. R., & HEGARTY, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition*, **27**, 741-750.

RUDDLE, R. A., & LESSELS, S. (2006). For efficient navigational search, humans require full physical movement, but not a rich visual scene. *Psychological Science*, **17**, 460-465.

SHELTON, A. L., & MCNAMARA, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review*, **4**, 102-106.

SHELTON, A. L., & MCNAMARA, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology*, **43**, 274-310.

VALIQUETTE, C. M., MCNAMARA, T. P., & LABRECQUE, J. S. (2007). Biased representations of the spatial structure of navigable environments. *Psychological Research*, **71**, 288-297.

WERNER, S., & SCHMIDT, K. (1999). Environmental reference systems for large-scale spaces. *Spatial Cognition & Computation*, **1**, 447-473.

WILLIAMS, B., NARASIMHAM, G., WESTERMAN, C., RIESER, J., & BODENHEIMER, B. (2007). Functional similarities in spatial representations between real and virtual environments. *ACM Transactions on Applied Perception*, **4** (2, No. 12).

NOTE

1. Latency did not vary as a function of imagined perspective and was slightly, but nonsignificantly, greater for the 0º–135º than for the 135º–0º learning order.

(Manuscript received July 11, 2007;
revision accepted for publication September 6, 2007.)