The Integration of Spatial Information across Different Perspectives

Jan M. Wiener (mail@jan-wiener.net) [1,2]
Collège de France, Laboratoire de Physiologie de la Perception et de l'Action (LPPA)
11, place Marcelin Berthelot, 75005 Paris, France

Tobias Meilinger (tobias.meilinger@tuebingen.mpg.de) [1]
Max-Planck-Institute for Biological Cybernetics
Spemannstr. 38, 72076 Tübingen, Germany

Alain Berthoz (alain.berthoz@college-de-france.fr)
Collège de France, Laboratoire de Physiologie de la Perception et de l'Action (LPPA)
11, place Marcelin Berthelot, 75005 Paris, France

Abstract

The integration of spatial information across different perspectives or viewpoints is a frequent spatial task, yet relatively little is known about it. In the present study, participants were shown three target locations from one point of view. After walking away, they either returned to the same location or to a novel location before being presented with three additional target locations. Their task was to plan and navigate the shortest possible path to visit all six target locations. To successfully solve the task, participants had to integrate different pieces of spatial information acquired from two viewpoints. We measured errors and the time to reach the first target. An increased number of errors in the condition including a perspective shift strongly suggest that participants encoded different views which had to be aligned in order to be integrated. The fact that the increase in errors primarily originated from the target locations presented first, indicates that the first view was transformed into the perspective of the second view. Neither egocentric updating, allocentric orientation-independent memory, nor allocentric reference axis theory can explain these results.

Keywords: Spatial memory; short term memory; integration; perspective; mental rotation; view-dependent; allocentric; egocentric; spatial planning

Introduction

The world around us can not be perceived in one glance only. Many tasks in our daily life require the integration of spatial information perceived at different times and from different viewpoints into a coherent, unified representation. In the last couple of years progress has been made in examining the integration of information across time (e.g., Brockmole, Irwin & Wang, 2003; Kurmar & Jiang, 2005). Several studies also examined spatial memory acquired from different points of view (e.g., Christou & Bülthoff, 1999; Shelton & McNamara, 2001). The experimental tasks in these latter studies (i.e., recognition or judgments of relative direction), however, did not require integrating separate information acquired from different viewpoints. Hence, up to now, it is largely unknown how spatial information perceived from different viewpoints and at different times is integrated. The present study aims at closing this gap.

Although the integration of spatial information across viewpoints has not been examined intensively, existing theories of spatial memory and processing provide hypotheses of how this is accomplished. In the following we briefly sketch these theories and work out their predictions for the case of spatial integration. In the experiment these predictions are then tested. The competing theories of spatial memory are: (1) egocentric updating, (2) view dependent memory, (3) allocentric reference axes and (4) allocentric orientation-independent memory.

Egocentric updating. In egocentric representations the angle, distance, and orientation of each object is represented relative to the body orientation of the navigator (cf. Burgess, 2006; Klatzky, 1998; Wang & Spelke, 2002). For example, an object is five meters in front of me slightly to the right. This is best described by a vector (plus an angle representing the orientation of the object if necessary). During navigation one keeps track of the objects' positions by updating these vectors based on perceived ego-motion information, even if the objects are not visible (spatial updating). Increasing evidence shows that such egocentric vector representations are rather transient (Waller & Hodgeson, 2006). When representing spatial information from two different perspectives, positions of objects perceived from the first viewpoint have to be updated while moving to the second viewpoint. Objects perceived from the second viewpoint are then represented by adding more vectors. In order to solve spatial tasks requiring information perceived from both viewpoints all vectors are used. Movement, especially turns, will decrease updating performance (cf., Klatzky, Loomis, Beall, Chance & Golledge, 1998, Riecke & Wiener, in press).

View dependent memory. View-dependent memory of an environment predicts that memory performance is better when being aligned with the perspective the environment was originally experienced from (e.g., Christou & Bülthoff,

---

[1] J.M.W. and T.M. contributed equally to this work. [2] Now at: Centre for Cognitive Science, University of Freiburg, Germany

1999; Mallot & Gillner, 2000; Wang & Spelke, 2002). If misaligned, compensatory transformations are necessary and the memory performance depends on the magnitude of misalignment. Perceiving spatial information from different perspectives will lead to the representation of multiple views. In order to integrate them, they have to be aligned and super-imposed. Therefore, one of the views has to be transformed either by mental rotation or, if possible, by updating. View dependent memory differs in two crucial aspects from egocentric updating. First, views can be stored in memory, accessed afterwards, and transformed offline. They can be remembered and mentally rotated into different perspective. Egocentric updating, in contrast, requires continuous updating during navigation. Second, although view-dependent memory has been suggested to consist of egocentric relations (e.g., Burgess, 2006; Wang & Spelke, 2002), views are not necessarily limited to them. Within an encoded view also allocentric or object-to-object relations can be represented, i.e., the view can be represented as a whole including allocentric relations between parts of the view (cf. Meilinger, Riecke & Bülthoff, 2007). Despite this potential allocentric nature, spatial relations are still better accessible from the experienced perspective.

Allocentric reference axes. Encoding spatial information relative to one, or even more, reference axes is also a form of allocentric encoding (e.g., Shelton & McNamara, 2001; Rump & McNamara, 2007). According to this theory information from any perspective is encoded relative to an environmental reference axis. Performance in tasks such as judgments of relative direction is best when being aligned with the reference axes. The orientation of a reference axis originates either from the initial contact with the environment or it coincides with the main orientation(s) of the environment (e.g., parallel to the longer walls of a rectangular room). According to this theory, memory performance in a spatial integration task mainly depends on being aligned with the reference direction during retrieval. It is independent from which perspectives or viewpoints the information was experienced and encoded.

Allocentric orientation-independent memory. The last theoretic position states that spatial information is represented merely allocentrically, i.e., in object-to-object relations (e.g., Burgess, 2006; Sholl, 2001; Holmes & Sholl, 2005). No performance advantage for being aligned with an experienced view or a reference axis is assumed. The memory is thought to be orientation or viewpoint-independent. Accordingly, spatial information perceived from different viewpoints is encoded in the same orientation-independent representation and is equally well accessible from any other viewpoint.

We described four theoretic positions which can explain integrating spatial information from different perspectives. Note, however, that these positions do not necessarily exclude each other, i.e. participants might represent their environment in multiple ways at the same time. This is almost always assumed with allocentric representations which are always thought to exist in addition to an egocentric one (e.g., Burgess, 2006; Sholl, 2001; Rump & McNamara, 2007). Our experiment tested which of these representations is involved in integrating spatial information across different perspectives.

Methods

Participants

Eight females and eight males between the ages of 24 and 37 (*M* = 28 years, *SD* = 3.4 years) participated in the experiment. Most of them worked in the Collège de France and were rewarded with chocolate for their participation.

Material

For the experiment the participants used the “magic carpet” (see Figure 1). It consists of 13 pressure sensitive tiles (30x30cm) that can be illuminated individually and that are embedded in a blue carpet. A computer controlled the illumination (i.e., stimulus presentation) as well as data recordings. The entire magic carpet was surrounded by a dashed line. In the center of each of the four sides of the carpet a start location was marked at the height of the dashed surrounding line (see Figure 1 and 2).

Procedure

The experimental task and the general procedure were as follows: Participants stood at one of the four start locations facing the magic carpet when 3 of the 13 tiles briefly lit up. Subsequently they turned to the right and walked along the dashed line to the next corner. Depending on the condition, participants either turned around and walked back to the start location (same perspective condition) or they continued to the next start location (different perspective condition) where they were presented with 3 additional target tiles. Their task was to remember all 6 target tiles and then step onto these tiles in an order that minimizes overall distance of the chosen path. The task of planning the shortest path required participants to integrate the target tiles perceived from both viewpoints.

The exact timing of a single trial was as follows. Participants stood at an X in front of the carpet looking down on their shoes. A cap prevented them from seeing the carpet with peripheral vision. After a beep they lifted their head and were given 2000 ms time to orient themselves before the first three target tiles lit up for 1000 ms. After this presentation participants looked at the carpet for an additional 1500 ms to encode the locations of the target tiles. The second beep indicated the end of this period. Participants lowered their head, turned 90° to the right, and walked along the dashed line to the next corner (see Figures 1 and 2). Here they either turned left around 180° and walked back to the start position (same perspective condition) or they turned 90° to the left and walked to the next X (start location) in front of them (different perspective condition, see Figure 2). After reaching the X participants turned towards the carpet while still looking at their shoes as they did the whole time during walking. 6500 ms after the last beep another beep indicated that they could lift their head, look at the carpet, and orient themselves. After 2000 ms three more target tiles lit up for 1000 ms. These were always different tiles than the first three tiles. A last beep indicated the end of the second presentation and participants were allowed to start walking. They were instructed to step only on the six target tiles in such an order that the overall path length was minimal.

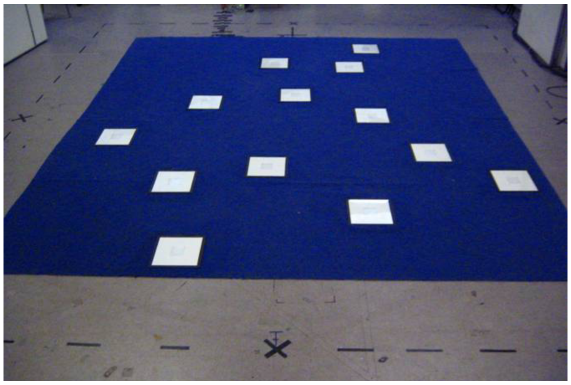

Figure 1: The magic carpet. The Xs mark view/start locations. The dashed lines indicate the routes used during walking.

We recorded participants' trajectories (i.e., the order which they stepped onto the tiles) and computed the number of errors (i.e., the number of tiles visited although they had not been illuminated during target presentation). To obtain a measure for planning performance we related the length of the chosen path to the shortest possible (optimal) path. Planning performance is expressed as percentage above optimal (PAO). A path with 10 PAO is 10% longer than the optimal solution. Furthermore, we calculated the start time (i.e., the time from the start signal until participants stepped on the first tile). Values deviating more than three standard deviations from the overall mean were replaced by the most extreme value observed within this interval.

The experiment consisted of 24 trials, subdivided into two blocks of 12 trials. One block comprised the same perspective condition, while the other block comprised different perspective condition. The order of conditions was balanced between participants. For each trial a different subset of 6 target tiles (2x3 tiles) was chosen.

In addition to varying the integration condition, the start location of the presentation, and therefore, also the start location for walking the planned path was varied. In the same perspective condition both these locations were identical. In the different perspective condition the starting location for walking the route was 90° rotated counterclockwise around the carpet, as participants walked there between presentations (see Figure 2). All participants began with the same starting location for the presentations. Every three trials they changed to the start location moving 90° counterclockwise, i.e., participants started three trials from the same start location before changing the start location. After one experimental block (12 trials) they were back at the initial start location. By these means all 4 start locations occurred equally often within each condition.

To familiarize participants with the timing and to reduce learning effects, they underwent 12 training trials. Participants were presented with three target tiles; they walked to the next corner to their right, went on to the next starting location or came back to the first starting location. Instead of being presented three more target tiles they had to step on the three target tiles in any order. The timing was identical to the first part of the presentation in the main experiment. After this training phase, participants were given additional training trials with the exact procedure of the test phase until they understood the rather complicated sequence of events in the test trials. After completing the experiment, participants were asked to describe how they solved the task. Overall, the whole experiment lasted about 1 hour.

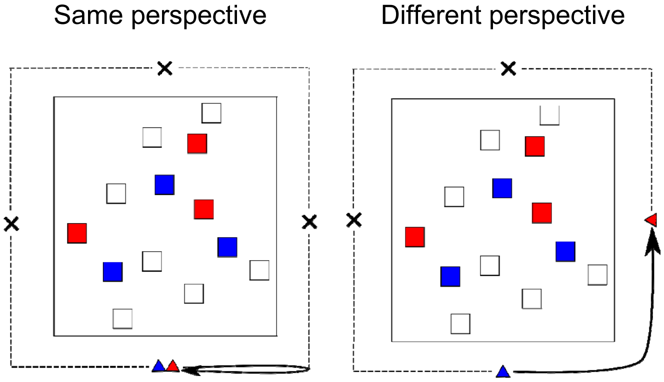

Figure 2: Schematic drawing of a trial. Participants saw the first three tiles (pained dark blue) from a location indicated here by the dark blue triangle. Then they turned right and walked to the corner before either returning to the same location (left side) or they walked on around the carpet (right side). From this location indicated by the bright red triangle they watched the second three tiles (painted bright red) before trying to walk the shortest route across these six tiles.

Predictions

The four theories of spatial memory and processing described above predict different outcomes for the experiment:

According to the **egocentric updating** the vectors to the three target tiles presented first have to be updated during the movement between presentations and are not stored in memory. It is crucial to emphasize that such updating is required independent of whether or not participants come back to the viewpoint from which they perceived the first presentation. Updating is known to be vulnerable to movements, especially to turns (cf. Klatzky et al., 1998; Riecke & Wiener, in press), larger turns should therefore result in more errors. In the same perspective condition participants turned 360° altogether as compared to 270° in the different perspective condition. If participants encoded the target locations purely egocentrically they should show better memory performance in the different perspective condition than in the same perspective condition. It is, however, possible that participants used the experimental room as global compass information to correct for accumulating errors during rotations. In that case, the actual turning angle becomes irrelevant and similar performance between conditions is expected.

The **view-dependent memory** position predicts that participants encode two independent views during the presentations. In order to solve the planning task, these views have to align to be integrated. In the same perspective condition the two views are aligned anyway. In the different perspective condition, in contrast, one view has to be rotated to align with the other view – either the first view is rotated towards the second view or vice versa. In any case this additional mental process should result in performance decrease as compared to the same perspective condition.

The **allocentric reference axes** theory and the **allocentric orientation-independent** theory make similar predictions with respect to the experimental conditions: in these theories spatial locations are either encoded with respect to salient axes in the environment (reference axes theory) or with respect to other locations in the environment (allocentric orientation-independent theory). The location and orientation of the observers (i.e., their viewpoint) during encoding have no influence on the representation itself. Therefore, both of these theories predict no differences in memory and integration performance between conditions.

The reference axes theory, however, predicts that participants should perform better if aligned with the axis used to encode the target locations. This axis should either coincide with the orientation they first experienced the room from or with the main orientation of the rectangular room (long side).

Results

The results revealed neither an order effect with respect to conditions (errors: $t(14) = .83$, $p = .419$, $d' = .42$; start time: $t(14) = .38$, $p = .710$, $d' = .19$; PAO: $t(14) = 1.67$, $p = .118$, $d' = .83$), nor a gender effect (errors: $t(14) = .96$, $p = .352$, $d' = .48$; start time: $t(14) = 1.60$, $p = .132$, $d' = .80$; PAO: $t(14) = 0.12$, $p = .908$, $d' = .06$). For the further analysis only the collapsed data was used.

On average, participants made 0.79 errors per trial, that is, they stepped on 0.79 tiles that were not illuminated and therefore no target tiles for the respective trial. The error rate differed between experimental conditions (same perspective – different perspective, see Figure 3). Specifically, participants made more errors in the different perspective condition than in the same perspective condition ($t(15) = 4.30$, $p = .001$, $d' = 1.1$).

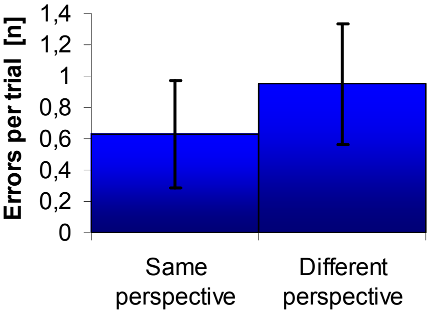

Figure 3: Differences in errors due to perspective change. Means and standard deviations are displayed.

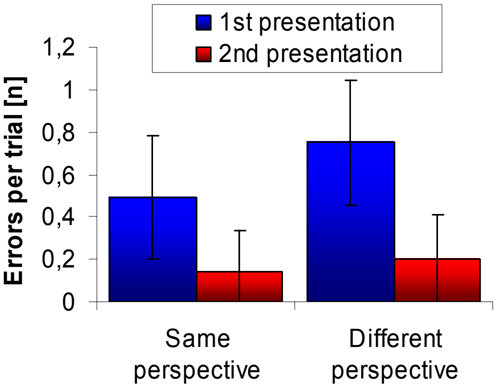

Figure 4: Differences in errors due to perspective shift for the first part and the second part of the presentation. Means and standard deviations are displayed.

A more detailed analysis of the error pattern reveals two further effects: (1) Participants more often missed target tiles that were presented in the first part of the presentation than in the second part (see Figure 4; $t(15) = 6.34$, $p < .001$, $d' = 1.58$); (2) the main difference in error rate between the conditions results from the target tiles presented first (see Figure 3; $t(15) = 3.88$, $p = .001$, $d' = 0.97$). In contrast, no difference was found for the error rate of the target tiles presented second $t(15) = 1.04$, $p = .316$, $d' = 0.26$).

An analysis of participants' start time demonstrates that the reported differences in error rate cannot be explained by a speed-accuracy trade-off. No significant differences were found between conditions– the numerical difference is even parallel to the errors (different perspective condition: 3.51s; same perspective condition: 3.18s; $t(15) = 2.07$, $p = .057$, $d' = 0.52$). Participants' path planning performance did not differ between conditions (PAO in the different perspective condition: 5.3%; PAO in the same perspective condition: 6.0%; $t(15) = 1.27$, $p = .22$, $d' = 0.32$).

An analysis of participants' performance in dependence of the start place revealed that participants were faster when starting from a location aligned with the main orientation of the experimental room ($F(3,45) = 2.81$, $p = .050$, $\eta^2 = .16$; errors: $F(3,45) < 1$). They also planned shorter routes (PAO: $F(3, 45) = 3.09$, $p = .036$, $\eta^2 = .17$). This effect was independent of the experimental condition. No effects were found for the viewpoint from which the first three tiles were presented (errors: $F(3, 45) < 1$; time: $F(3, 45) = 1.03$, $p = .387$, $\eta^2 = .06$; $F(3, 45) = 2.09$, $p = .115$, $\eta^2 = .12$).

Discussion

In this study we examined the integration of spatial information across different perspectives (i.e., across different viewpoints). Contrary to temporal integration of visuo-spatial stimuli, little is known about spatial integration. Therefore, we derived and tested hypotheses about spatial integration from existing theories of spatial memory and processing. In the experiment, participants were shown two subsequent presentations of three target places that had to be integrated. In the different perspective condition, participants moved to a new viewpoint between presentations, while they viewed both presentations from the same viewpoint in the same perspective condition. Results demonstrate that performance was better in the same perspective condition than in the different perspective condition.

This result was predicted by the view-dependent memory only (see Predictions), stating that spatial layouts are encoded in the orientation they were experienced (cf., Christou & Bülthoff, 1999; Mallot & Gillner, 2000; Meilinger et al., 2007). If relevant spatial information is perceived from one viewpoint only (same perspective condition), encoded views are already aligned and can easily be integrated. If, however, spatial information is perceived from different viewpoints (different perspective condition) the different pieces of spatial information have to be transformed into the same coordinate system. In other words, views encoded from different viewpoints have to be aligned before being integrated, e.g., by mentally rotating one view to match the other. This additional spatial transformation then leads to an increased error rate as observed in the different perspective condition. View dependent memory has been suggested to also include allocentric relations between parts or objects captured in the view (Meilinger, Riecke & Bülthoff, 2007). While this is not necessary to explain our results so far, informal reports strongly suggest that participants memorized the target tiles in patterns such as lines or triangles. Essentially, this constitutes an allocentric strategy as relations between objects are encoded.

What do the other theories propose? Egocentric updating predicts that errors should increase with increasing distance and turning angle covered between subsequent presentations (cf., Burgess, 2006; Klatzky, 1998; Wang & Spelke, 2002). In the same perspective condition, participants walked the same distance between presentations, but the overall turning angle during navigation was larger than in the different perspective condition (see Procedure Section). Accordingly, the egocentric updating theory predicted that performance would be better in the different perspective condition than in the same perspective condition. Egocentric updating, therefore, does not account for the results. Note that the updating of the target positions during movements is an inevitable feature of this theoretic position. Pure egocentric updating does not allow storing views and accessing them later (this is predicted by the view dependent theory only). Even if the experimental room was used as a global or compass reference during egocentric updating to correct for errors in estimated turning angle, the results cannot be explained: in this case no differences between experimental conditions were predicted.

Similar, neither allocentric orientation-independent theory (e.g., Burgess, 2006; Sholl, 2001; Holmes & Sholl, 2005) nor reference axes theory (e.g., Shelton & McNamara, 2001; Rump & McNamara, 2007) can explain the results. Allocentric orientation-independent representations would not predict any performance difference due to perspective changes, as all spatial relations are encoded perspective- and orientation-independent anyway. Reference axes theory would predict an advantage when being aligned with the main reference axis of the environment. However, as all starting locations and, therefore, all orientations with respect to the room occurred equally often in both conditions, the effect can not be explained by reference axis theory.

Independent of the specific perspective shift condition, reference axes theory predicts that participants perform better when being aligned with a reference axis. This prediction was supported: Participants were faster when starting along the main orientation of the room (i.e., the direction of the room's longest extension) and they walked shorter routes. Note, however, that this orientation was also the orientation participants were most familiar with due to prior experience with the experimental room. Future experiments have to show whether better performance was due to encoding the environment along a reference axis (in addition to encoding multiple views of it) or whether this particular view was just easier to process for the participants due a higher familiarity.

We also found better memory performance for tiles presented in the second presentation. This effect is also found in studies examining temporal integration (e.g., Brockmole et al., 2003, Kurmar & Jiang, 2005) and can be attributed to the shorter retention interval for the second presentation.

How exactly does spatial integration across different perspectives work? Our results indicate that participants encoded separate views perceived from two different perspectives. These views had to be aligned and integrated to solve the path planning task. Such an alignment process could, for example, be achieved by mental rotations and is possible in two directions: either the first experienced view is rotated towards the second view or vice versa. Analyzing the error pattern in more detail indicates that the former is the case: the higher number of errors in the different perspective condition as compared to the same perspective condition mainly originates from the target tiles that were presented first.[2] If one assumes that a process such as mental rotation decreases performance (e.g. Shepard & Metzler, 1971), this result indicates that the first view had been mentally rotated to align it with the second view. If the second view would have been rotated to align it with the first view, an increase in errors for the second view, but not the first view would have been expected. Note that aligning the first view with the second view could also be achieved by the updating of the entire view during walking. Assuming a view dependent memory such an updating would, of course, only be necessary in the different perspective condition. In the same perspective condition, participants could rely on memorized views.

Why do participants rotate the first view to align it with the second view and not vice versa? The most obvious and parsimonious explanation is that participants transform the relevant information into the perspective in which they want to use it (i.e., in which they plan their route and start walking). Results from a second experiment that is not reported here support this hypothesis. Rotating the first view to align with the second view not only minimizes mental effort, but also allows for immediate action.

Conclusion

Despite the fact that the integration of spatial information across different perspectives or viewpoints is a frequent spatial task, relatively little is known about it. In this work, we investigated the underlying mechanisms, strategies, and representational formats. The results are explained best when assuming a view dependent encoding of spatial information. Neither purely egocentric, nor allocentric theories assuming an orientation-independent spatial memory or a reference axis could explain this result.

Taken together, our results suggest that participants stored independent views of the two presentations of target tiles. In order to integrate them, they then transformed the first view to align and superimpose it with the second view.

Acknowledgments

This research was supported by the EU grant "Wayfinding" (6th FP - NEST) and the VW-Foundation (Tandem Project: Wayfinding strategies in Behavior and Language). The authors thank the participants for their help.

References

Brockmole, J. R., Irwin, D. E., & Wang, R. F. (2003). The locus of spatial attention during the temporal integration of visual memories and visual percepts. *Psychonomic Bulletin & Review, 10,* 510-515.

Burgess, N. (2006). Spatial memory, how egocentric and allocentric combine. *Trends in Cognitive Science, 10,* 551-556.

Christou, C.G. & Bülthoff, H.H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition, 27,* 996-1007.

Engelkamp, J. & Zimmer, H.D. (1984). Motor programme information as a separable memory unit. *Psychological Research, 46,* 283.

Holmes, M.C. & Sholl, M.J. (2005). Allocentric coding of object-to-object relations in overlearned and novel environments. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 1069-1078.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition - An interdisciplinary approach to representation and processing of spatial knowledge* (pp. 1-17). Berlin: Springer.

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science, 9,* 293-298.

Kurmar, A. & Jiang, Y. (2005). Visual short-term memory for sequential arrays. *Memory & Cognition, 33,* 488-498.

Mallot, H.A. & Gillner, S. (2000). Route navigation without place recognition: What is recognized in recognition-triggered responses? *Perception, 29,* 43-55.

Meilinger, T., Knauff, M., & Bülthoff, H. H. (in press). Working memory in wayfinding - a dual task experiment in a virtual city. *Cognitive Science.*

Meilinger, T., Riecke, B.E. & Bülthoff, H.H. (2007). Orientation Specificity in Long-Term-Memory for Environmental Spaces. *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 479-484).

Riecke, B.E. & Wiener, J.M. (2007). Can People Not Tell Left from Right in VR? Point-to-origin Studies Revealed Qualitative Errors in Visual Path Integration. *Proceedings of IEEE Virtual Reality 2007, (pp. 3-10).*

Rump, B. & McNamara, T.P. (2007). Updating Models of Spatial Memory. In T.Barkoswky, M.Knauff, G. Ligozat, & D.R. Monello (Eds.), *Spatial Cognition V* (pp. 249-269). Berlin: Springer.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43,* 274–310.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of threedimensional objects. *Science, 171,* 701-703.

Sholl, M.J. (2001). The Role of a Self-Reference System in Spatial Navigation. In D.R. Montello (Ed.), *COSIT 2001* (pp. 217-232). Berlin: Springer.

Waller, D. & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 867-82.

Wang, R.W. & Spelke, E.S. (2002). Human spatial representations: insights from animals. *Trends in Cognitive Sciences, 6,* 376-382.

---

[2] The error rates in the second presentation (.17 errors/trial) were high enough to rule out a bottom-effect which could prevent dicovering existing effects.