

The integration of spatial information across different viewpoints

Tobias Meilinger · Alain Berthoz · Jan M. Wiener

© Psychonomic Society, Inc. 2011

Abstract The integration of spatial information perceived from different viewpoints is a frequent, yet largely unexplored, cognitive ability. In two experiments, participants saw two presentations, each consisting of three targets—that is, illuminated tiles on the floor—before walking the shortest possible path across all targets. In [Experiment 1](#), participants viewed the targets either from the same viewpoint or from different viewpoints. Errors in recalling targets increased if participants changed their viewpoints between presentations, suggesting that memory acquired from different viewpoints had to be aligned for integration. Furthermore, the error pattern indicates that memory for the first presentation was transformed into the reference frame of the second presentation. In [Experiment 2](#), we examined whether this transformation occurred because new information was integrated already during encoding or because memorized information was integrated when required. Results suggest that the latter is the case. This might serve as a strategy for avoiding additional alignments.

Keywords Spatial memory · Integration · Alignment · Short term memory · Reference frame · Spatial planning

Introduction

The world around us cannot be perceived in only one glance. Many tasks in our daily life require the integration of spatial information experienced at different times and from different viewpoints. Consider the many glances it requires to see all of the locations in your home: You are nevertheless able to integrate them. This allows you to directly approach your fridge, your bedside lamp, your garbage can, and so on, from almost any location—even with your eyes closed. Or imagine finding shortcuts, planning novel routes, or pointing to distant locations within a city recently explored. These tasks require the integration of information that has been acquired at different times and from different viewpoints. Although many earlier studies have demonstrated that humans are indeed capable of integrating separate spatial representations (e.g., Golledge, Ruggles, Pellegrino, & Gale, 1993; Holding & Holding, 1989; Ishikawa & Montello, 2006; Maguire, Burke, Phillips, & Staunton, 1996; Moar & Carleton, 1982; Montello & Pick, 1993), very little is known about *how* they do so.

In the present study, we defined spatial integration as the process of combining different spatial representations that have been formed by multiple experiences within a single frame of reference or coordinate system (see also Greenauer & Waller, 2010; McNamara, Slucenski, & Rump, 2008; Meilinger, 2008). In this study, we particularly focused on two questions: First, do costs that arise during integration relate to the process of aligning differently oriented spatial representations? Second, which reference frame is used for

Tobias Meilinger and Jan M. Wiener contributed equally to this work.

T. Meilinger (✉)
Max-Planck-Institute for Biological Cybernetics,
Spemannstr. 44,
72076 Tübingen, Germany
e-mail: tobias.meilinger@tuebingen.mpg.de

T. Meilinger · A. Berthoz · J. M. Wiener
Collège de France,
11, place Marcelin Berthelot,
75231 Paris Cedex 05, France

J. M. Wiener (✉)
Bournemouth University,
Poole House, Talbot Campus, Fern Barrow Poole,
Dorset BH12 5BB, UK
e-mail: jwiener@bournemouth.ac.uk

integrating spatial information that was acquired from different viewpoints? By this, in the present work, we aimed to extend earlier work on spatial memory to conditions in which learning occurs across multiple perspectives, as in most natural learning situations.

Although integration costs such as decreased memory performance have been reported for some spatial tasks (Golledge et al., 1993; Hanley & Levine, 1983; Montello & Pick, 1993; Yamamoto & Shelton, 2008), researchers in other studies have not observed these costs (Holding & Holding, 1989; Moar & Carleton, 1982; Maguire et al., 1996). One crucial difference between these two classes of studies was the time at which integration happened. If the experimental paradigm allowed participants to integrate different representations during or shortly after encoding, no integration costs were observed during the recall of the spatial information. On the other hand, if the paradigm did not allow participants to integrate during encoding but integration happened when spatial information was required, integration costs were observed. Yamamoto and Shelton (2008), for example, asked participants to learn two arrays of objects presented in successive presentations. After being led into another room, participants were asked to conduct an imagined pointing task (a judgment of relative direction). Pointing within an array yielded better performance than did pointing between arrays demonstrating integration costs. However, when given sufficient time to integrate the two arrays immediately after presentation—that is, during encoding—no differences were observed between within and between array pointings.

Hanley and Levine (1983) asked participants to haptically explore two simple two-leg paths laid out on a table in front of them. After learning each path separately, participants were informed about where the paths overlapped and were then asked to draw lines between locations on the paths. Participants performed better when drawing lines between locations on the same path than when drawing lines between locations on different paths, again demonstrating integration costs. This effect, however, vanished when granting additional learning time with each separate path after informing participants about the connection point between the paths. In several other studies, researchers tested the integration of routes through navigational spaces, reporting more precise direction estimates between locations on the same route than on different routes (Golledge et al., 1993; Montello & Pick, 1993). Between-route direction estimates improved with additional training after participants were informed about the interconnection between routes (Ishikawa & Montello, 2006). No integration costs were observed after repeated video or slide presentations of simple routes where integration may have happened during encoding (Holding & Holding, 1989; Maguire et al., 1996; Moar & Carleton, 1982). These studies reveal specific costs of integrating

separately encoded spatial information during recall. With sufficient training, and after being informed about the interrelation between different pieces of spatial information, these costs can be overcome, probably by already integrating information during learning.

What causes integration costs as described above? One possible source is a transformation process in order to align different representations into the same reference frame before combining them. In the present experiments, we examined whether such transformation costs contribute to the costs associated with spatial integration. Note that a prerequisite for transformation costs is that spatial integration indeed happens within a single spatial reference frame. Recent evidence by Greenauer and Waller (2010) supports this assumption. Greenauer and Waller asked participants to learn two overlapping object arrays that were presented successively from the same viewpoint. The intrinsic orientation of both arrays (i.e., their symmetry axes) was different and determined the orientations along which participants performed best when pointing to within-array targets, demonstrating that participants formed separate representations. However, when asked to point between objects of different arrays, the separate representations had to be integrated. In this case, participants performed best when mentally oriented along a single common orientation, strongly suggesting that integration happened within a single reference frame. The orientation of this reference frame did not necessarily coincide with the intrinsic orientation of one of the two arrays and originated from the learning perspective or from the intrinsic orientation of the overall array. Although Greenauer and Waller's study elegantly demonstrated that separate spatial representations are integrated within a single reference frame, the associated costs were not addressed. It is conceivable that integration costs arise from transforming spatial information from differently oriented reference frames into a common reference frame (or subsuming them under a higher-level reference frame). This assertion, however, awaits empirical validation. In the present study, we compared the integration of two aligned spatial representations—encoded from the same viewpoint—with the integration of two misaligned representations that were encoded from different viewpoints. Any differences between these two situations would indicate that aligning differently oriented reference frames into a common reference frame that coincides with one of the original reference frames contributes to integration costs.¹

¹ Note that spatial integration encompasses all processes required to use two separately formed spatial representations. In addition to alignment, other processes might contribute. For example, separate spatial representations might be fused into a single representation. Whether such processes also contribute to the costs associated with spatial integration is beyond the scope of the present study.

The second question addressed in the present study concerned the selection of the reference frame that is used when integrating spatial information that was acquired from different viewpoints. In general, a single spatial layout experienced from one viewpoint or from different viewpoints is represented within a common reference frame. The orientation of this reference frame may be determined by the viewpoint (perspective) of the learner, by the intrinsic orientation of the spatial layout itself, or by the form of the surrounding space (Christou & Bühlhoff, 1999; Kelly & McNamara, 2008; Mou & McNamara, 2002; Shelton & McNamara, 2001). It is, however, an open question as to which reference frame is used when multiple pieces of spatial information that were experienced separately from different viewpoints are integrated. Kelly, Avraamides, and McNamara (2010) tested a similar situation; however, this did not require integration. Specifically, Kelly and colleagues examined whether a reference frame established when encoding an array of objects from one viewpoint was also used to encode an additional overlapping array that was experienced either from the same or from a different viewpoint. When both arrays were encoded from the same viewpoint, participants' pointing performance (judgments of relative direction between objects of one array) was best when aligned with that viewpoint. However, when the arrays were encoded from different viewpoints, participants performed best when aligned with the viewpoint from which the first array had been encoded or when oriented orthogonal to that viewpoint. This result suggests that the orientation in which the first array was encoded was also used to encode the second array. This holds true, even if the two arrays were encoded using different sensory modalities—that is, if one array was explored haptically while the other array was explored visually. Note that participants were not required to point between objects from the different arrays and, therefore, actual integration was not required. Nevertheless, arrays experienced from different viewpoints were encoded within the same frame of reference. The results of the present study demonstrate that reference frames selected by prior experience can determine how novel information is encoded, even if the task itself does not require integrating the different pieces of information.

Integration does not necessarily happen within already established (intrinsic) reference frames. As was mentioned earlier, results by Greenauer and Waller (2010) demonstrate that integration can also happen in a reference frame that is determined by the intrinsic orientation of the overall array of objects, or by the perspective onto that overall array. Similarly, orientation-free reference frames have been proposed (Byrne, Becker, & Burgess, 2007; Presson, DeLange, & Hazelrigg, 1987). Integrating within such an orientation-free reference frame would require any spatial information to be transformed into that reference frame (Byrne et al., 2007).

In addition to integrating novel spatial information within already-established reference frames or within a reference frame unrelated to established ones, two other possibilities seem plausible. First, integration can happen in the reference frame in which novel information is encoded. For example, when navigating through environments, one might encode and integrate novel information from the point of view of the current perspective. This would require that information encoded earlier is transformed into the current perspective (e.g., by spatial updating as movements take place, or by mental rotation). Second, spatial information acquired at different times and from different viewpoints is kept separate until an integrated representation is required. In this case, all representations have to be transformed into the reference frame that is relevant for action.

The experiments in this study were designed (a) to investigate whether alignment processes contribute to integration costs, and (b) to investigate which reference frame is chosen for integration. If the integration of representations in fact yields alignment costs—in the present article, memory errors—we expect worse performance for misaligned representations than for aligned representations. Moreover, memory errors should also be predictive for the reference frame used for integration: If a later representation is transformed toward an earlier one, this should yield higher costs for items encoded later rather than earlier, and vice versa. If, on the other hand, a novel or an orientation-free reference frame was used for integration, all representations have to be transformed toward that reference frame, and no overall difference in memory accuracy should be observed for information learned during the different presentations (exceeding the decay through time). Table 1 sums up the assumptions guiding the experimentation.

Experiment 1

In **Experiment 1**, participants viewed two presentations of three target locations each (see Fig. 1). Between presentations, participants either moved and came back to the original location (same viewpoint condition) or they moved to a new location (different viewpoint condition). After the second presentation, participants were asked to plan and navigate the shortest path to visit all targets (for similar planning tasks, see Wiener, Ehbauer, & Mallot, 2009). Planning an efficient path requires integration (i.e., considering all locations together). The quality of the planned paths is thus an indicator of whether or not participants successfully integrated information. Memory errors are attributed to target locations presented during the first or the second presentation, allowing the identification of which representation was transformed.

Table 1 Assumptions guiding the present experiments

- (1) Spatial integration happens within a single reference frame
- (2) Transforming misaligned spatial representations into a common reference frame requires mental processing, which may result in memory errors or increased latencies relative to an appropriate baseline (i.e., integrating already-aligned representations).
- (3) Spatial information acquired from a novel viewpoint 2 can be integrated with existing spatial memory acquired from a viewpoint 1 using the following:
 - (a) the already-established reference frame corresponding to viewpoint 1 (cf. Kelly et al., 2010)
 - (b) the reference frame of viewpoint 2 from which the novel information is perceived
 - (c) the reference frame corresponding to the viewpoint at which the relevant spatial task is executed; in this case, two viewpoint-dependent representations are maintained in memory until the task is executed
 - (d) a reference frame abstracted from the views and corresponding to neither of them (e.g., an orientation-free reference frame, or a reference frame based on the dominant layout structure; cf. Greenauer & Waller, 2010)
- (4) If representation A is transformed to align with representation B, then errors will be higher on memory items encoded in representation A

Method

Participants Sixteen participants (eight females, eight males) took part in the experiment ($M = 28$ years, $SD = 3.4$ years). Most of them worked in the Collège de France and volunteered without incentive or payment.

Material/Apparatus The experimental setup, the “Magic Carpet” (developed at the LPPA by P. Leboucher and M. Ehrette) consisted of 13 tiles (30 x 30cm each) that could be illuminated individually and were equipped with pressure sensors. The tiles were embedded in a carpet (see Fig. 1). The illumination was used to highlight the target locations/tiles that participants had to remember. Each time they stepped on a tile, the pressure sensor recorded the passage of the participant. By these means, we obtained both time and trajectories (i.e., the straight-line connections between tiles). A custom-designed computer program controlled tile illumination as well as data recording. A dashed rectangle drawn on the floor surrounded the entire magic carpet. In the center of each of the four sides of the rectangle, a start location was marked by an “X” (see Figs. 1 and 2).

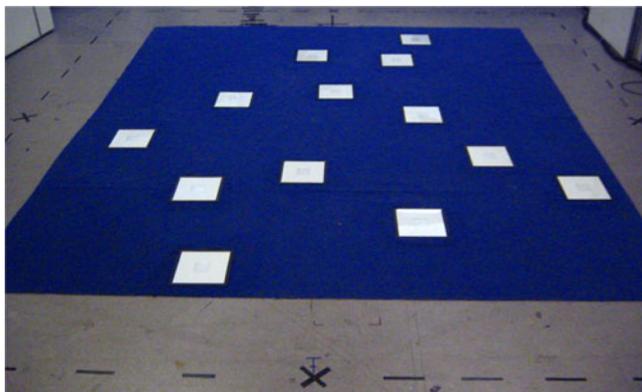


Fig. 1 A photograph of the magic carpet. The Xs mark view/start locations. The dashed lines indicate the paths used during walking

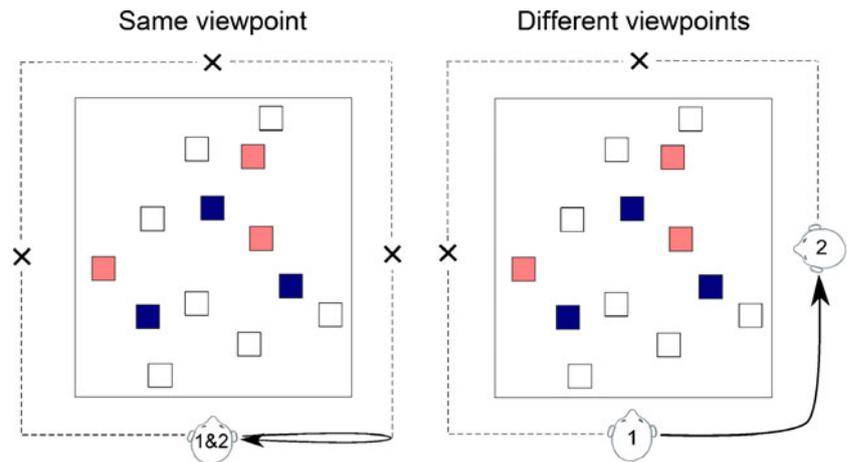
Procedure The experimental task was as follows: Participants stood at one of the start locations (see Fig. 1) oriented toward the carpet but looking down at their shoes. They wore a cap, which prevented them from seeing the carpet with peripheral vision and thus prevented visually updating the targets during later locomotion.

First presentation. Following a first beep, participants lifted their heads and were given 2,000 ms to orient themselves before the first three target tiles lit up for 1,000 ms. After target illumination was turned off, participants were allowed to look at the carpet for an additional 1,500 ms to encode the locations of the target tiles. According to experiments investigating spatial integration information across time, 1,500 ms is sufficient to encode information into short-term memory (Brockmole, Irwin, & Wang, 2003; Brockmole, Wang, & Irwin, 2002; Jiang, 2004). A second beep indicated the end of this period.

Walking. After the second beep, participants lowered their heads, turned 90° to the right, and walked along the dashed line to the next corner (see Figs. 1 and 2). Here, they either turned around 180° to the left and walked back to the original start position (same viewpoint condition), or they turned 90° to the left and walked to the next start location (different viewpoint condition). After reaching their destination, participants turned toward the carpet.

Second presentation. A third beep (6,500 ms after the second beep) indicated that participants could lift their heads, look at the carpet, and orient themselves. After 2,000 ms, three additional target tiles that were different from the first three tiles lit up for 1,000 ms. These were always different from the first three tiles. A last (fourth) beep indicated the end of the second presentation, and participants were allowed to start walking. They were instructed to step on all six target tiles in such an order that the overall path length was minimal. This planning task (cf. Wiener et al., 2009) required participants to integrate the six target locations that were perceived at different

Fig. 2 Schematic drawing of the procedure of a single trial. Participants saw the first three target tiles (dark blue) from “1.” Then, they turned right and walked to the corner before either returning to the same location (left side—same viewpoint condition) or they walked around the carpet (right side—different viewpoint condition). From this location “2,” they were presented with the three additional target tiles before planning the shortest route and walking across the six target tiles



times. Only if the target locations of the two presentations were considered simultaneously during path planning could optimal or near-optimal solutions be achieved (cf. Fig. 2).

To familiarize participants with the procedure and to reduce learning effects, they underwent 12 training trials in which they were required to remember three tiles only. In addition, participants received practice trials with the exact procedure of the test phase (usually two or three trials) until they understood the somewhat complicated procedure. Consequently, participants always knew beforehand where they were asked to go. After completing the experiment, participants were asked to describe how they solved the task. The entire experiment lasted approximately 45 min.

Design The design was within subjects, with trials blocked by condition and a counterbalanced order of conditions. The experiment consisted of 24 trials that were subdivided into two blocks of 12 trials each. One block comprised the same viewpoint condition, whereas the other block comprised the different viewpoint condition. The order of conditions was balanced between participants. A different subset of six target tiles (2 x 3 tiles) was chosen for each trial. The 24 subsets were presented in random order.

In addition to varying the viewpoints (same vs. different), we varied the start location. After three trials, the start location changed (90° counterclockwise). After 12 trials (one experimental block), participants were back at the initial start location. By these means, participants started from all four start locations equally often within each condition. This procedure allowed us to control for possible effects of reference frames intrinsic to the room or the carpet; that is, possible transformation angles toward any room-based reference frame used for integration were balanced between conditions.

Analyses Participants' trajectories were recorded, and the number of errors computed. An error was made if participants omitted a target tile—that is, if participants

did not step onto a tile that had been illuminated during target presentation. For trials without error, planning performance was analyzed to measure whether or not participants integrated information from multiple presentations. Planning performance is expressed as percentage above optimal (PAO), describing the deviation of the selected path from the optimal solution (i.e., shortest possible path that visits all six target tiles). A path with a PAO value of 10 is 10% longer than the optimal solution; a PAO value of 50 corresponds to a path 50% longer than the optimal solution (Wiener, Lafon, & Berthoz, 2008). If participants did not integrate the target tiles perceived at different times, they would have planned a path first visiting all tiles from one presentation (first or second) and then all tiles from the other presentation. Accordingly, we also calculated the PAO value for the best solution under the assumption that integration did not happen. Participants' path planning performance was then compared both to the overall optimal solution and to the best solution without integration. In this way, planning performance serves as a control for integration. Furthermore, we calculated start times (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 removed from the data. Errors and start time were analyzed with a linear mixed-model analysis (e.g., Snijders & Bosker, 1999), with the fixed factors of condition (same vs. different viewpoint) and presentation (first vs. second—for errors only), and with participants as the random factor. As compared with an ANOVA, this analysis is less restrictive with regard to distribution assumptions; it allows for varying effect sizes within different participants while still controlling for order of conditions and gender, which were included as fixed factors in the full factorial design (i.e., modeling all possible interactions). We report significant main effects for gender and order as well as significant interactions with condition and presentation.

Predictions As was outlined in the introduction, it was assumed that successful integration requires transforming separate spatial representations into a common reference frame (Greenauer & Waller, 2010; McNamara et al., 2008; Meilinger, 2008). **Experiment 1** was designed to address whether such transformations are associated with alignment costs. Depending on the nature of the reference frame chosen for integration, the predictions for **Experiment 1** were as follows (see also Table 1):

If integration takes place in a frame of reference that is independent of the viewpoints during learning (cf. Greenauer & Waller, 2010), all representations have to be transformed to align with that novel reference frame. In this case, no differences will be observed between the same viewpoint condition and the different viewpoint condition.

If, however, the frame of reference used for integration originates from one of the different representations (e.g., the first presentation: Kelly et al., 2010), then alignment costs will arise only in the different viewpoint condition, in which two separate viewpoint-dependent memories have to be aligned (i.e., a main effect condition is expected).

Increased alignment costs in the different viewpoint condition will also allow addressing the question of which reference frame was used for integration. If the second representation is transformed toward the first representation (cf. Kelly et al., 2010), higher costs for items encoded second are predicted. If, on the other hand, the first representation is aligned and integrated within the reference frame of the second representation, the opposite pattern is expected. Since we have to control for the overall effect of time (main effect presentation), we predict an interaction between condition and presentation. The same viewpoint condition therefore acts as a control for effects of time and movement.

Note that after tile locations are transformed and integrated, no differences in planning performance between conditions is expected. Path planning performance, therefore, is only an indicator of whether integration happened or not.

Results and discussion

Overall, participants showed good path planning performance: On average, the chosen paths were only 3.7% longer than the shortest possible path (see Fig. 3). For both groups, this performance was clearly better than predicted if no integration happened, $t(15) = 16.8$, $p < .001$.

The main aim of **Experiment 1** was to investigate whether spatial integration requires aligning spatial information that was acquired from different viewpoints. This is indicated in the main effect of condition in the linear mixed models analysis, $F(1, 740) = 15.1$, $p < .001$. Specifically, participants made more memory errors in the different

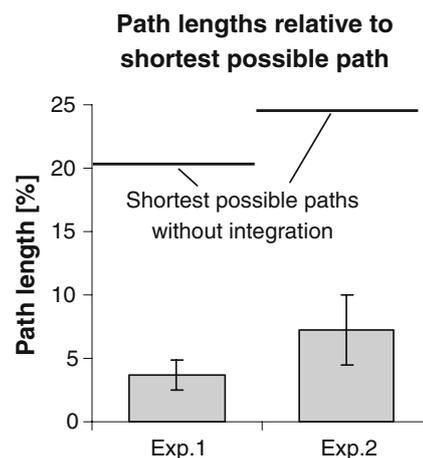


Fig. 3 Integration performance for both experiments. Perfect integration and perfect path planning would have resulted in no overshoot relative to the shortest possible path (i.e., 0% longer path length). No integration, but perfect planning would have resulted in 20% (**Experiment 1**) or in 25% (**Experiment 2**) longer paths, since participants would have walked first to all tiles from one presentation and then to the tiles from the other presentation. Means and 95% confidence intervals for the means are shown

viewpoint condition (errors per presentation and trial: $M = 0.47$; $SD = 0.68$) than in the same viewpoint condition ($M = 0.32$; $SD = 0.62$).² In the latter condition, all target tiles were perceived and encoded from the same viewpoint. The resulting representations were thus already aligned, and no further transformation costs resulting from alignment arose. A difference in memory performance between the conditions was predicted if the reference frame used for integration coincided with one of the reference frames during learning, but not if another reference frame—including an orientation-free reference frame—was utilized for integration. Thus, the results suggest that one viewpoint-dependent representation was transformed to align with the other viewpoint-dependent representation in the different viewpoint condition, which led to an increased error rate. We discuss the nature of this transformation in the **General discussion** section.³

The differences in error rates between experimental conditions cannot be explained by a speed–accuracy trade-off (i.e., differences in error rate do not originate from a corresponding difference in start times). In fact, start time was shorter in the same view point condition ($M = 3.11$ s; $SD = 1.29$ s) than in the different view point condition ($M = 3.36$ s; $SD = 1.39$ s), $F(1, 270) = 4.06$, $p = .045$.

² Note that the standard deviation includes variability between as well as within participants, although the applied tests used only within-participant variability.

³ Note that the observed 0.80 errors per trial (0.40 per presentation) correspond to 24.8% of the number of errors expected if participants randomly selected six out of the 13 tiles (chance level: 3.23 errors).

We also found a main effect of presentation, $F(1, 740) = 124, p < .001$. Specifically, error rate was higher for target tiles presented during the first presentation ($M = 0.62; SD = 0.77$) than for target tiles presented during the second presentation ($M = 0.17; SD = 0.40$). This finding is in line with earlier experiments investigating purely temporal integration of spatial information (Brockmole et al., 2002; Brockmole et al., 2003; Jiang, 2004) and could be explained by general memory decay over time. An alternative explanation is the use of different memory systems for the two presentations: Although locations from the first presentation had to be encoded into long-term memory and retrieved from long-term memory for the actual planning process, participants could rely on short-term memory for locations from the second presentation.

The main effect of condition is in line with data by Greenauer and Waller (2010), suggesting that differently oriented spatial representations—experienced from the same viewpoint—are integrated within a single reference frame. The present results extend these data to situations in which spatial information was experienced from different viewpoints. In addition, results suggest that alignment processes contribute to the costs observed during spatial integration. Such alignment costs might also explain spatial integration costs observed in various other studies (Golledge et al., 1993; Hanley & Levine, 1983; Montello & Pick, 1993; Yamamoto & Shelton, 2008).

The second aim of the present experiment was to examine which reference frame was used for integration. As was outlined in the predictions, the observed main effect of experimental conditions indicates that participants aligned and integrated separate representations in a reference frame that was aligned with one of the viewpoints during learning. In other words, either the representation encoded first was transformed toward the second, or vice versa. We did observe a significant interaction between the experimental conditions and the presentation, $F(1, 740) = 5.67, p = .018$, suggesting that mainly one of these transformations happened. The interaction was driven by a greater difference in error rate between conditions for the target tiles presented first, not second see Fig. 4; mean difference for first presentation, 0.26 errors ($SD = 0.76$), $F(1, 364) = 12.5, p < .001$; mean difference for second presentation, 0.06 errors ($SD = 0.39$), $F(1, 364) = 2.88, p = .091$. Here, the same viewpoint condition serves as a control for any alignment-unspecific effects resulting from presentation order, movement between presentations, and general memory decay. The increase in errors in the different viewpoint condition as compared with that in the same viewpoint condition thus primarily resulted from the target tiles presented during the first presentation, strongly suggesting that participants transformed the first representation to align with the second representation in the different viewpoint condition, not vice versa.

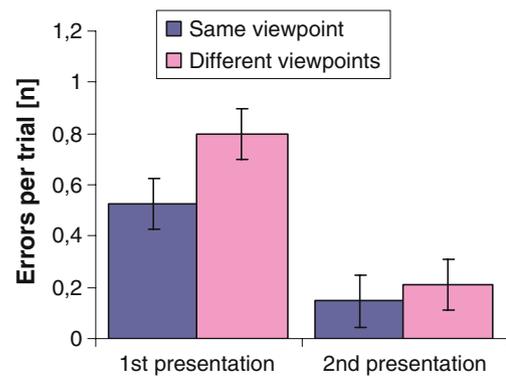


Fig. 4 Differences in error rate between conditions for tiles presented first and second. Means and 95% confidence intervals as estimated from the marginal means are shown

There was a significant three-way interaction among gender, presentation, and order of presentation in errors. This, however, did not change the main effect of presentation in any level combination of the interacting factors, $F(1, 740) = 10.3, p = .001$. No main effects of gender [start time: $F(1, 12) = 2.81, p = .120$], errors ($F < 1$), or of order (errors and time: $F < 1$) were observed.

Independent of the experimental conditions (same/different viewpoint), we also varied the start places. This was done as a control to ensure that any reference frame that was intrinsic to the room or to the carpet was aligned equally often with the four different start positions as well as with the experimental conditions. The alignment/misalignment of the start locations with an intrinsic axis of the experimental room/carpet did not influence error rate ($F < 1$) or response times, $F(3, 262) = 1.48, p = .219$. Such differences were expected if participants used reference frames for integration that were aligned with the intrinsic orientation of the room/carpet (Kelly & McNamara, 2008; Mou & McNamara, 2002; Shelton & McNamara, 2001; Valiquette & McNamara, 2007). We did not, however, test for oblique misalignments relative to the room/carpet (i.e., $45^\circ, 135^\circ$), which in these experiments often influence performance more than do orthogonal misalignments ($90^\circ, 180^\circ$).

In **Experiment 1**, we considered only alignment costs for the integration process itself. There may, however, have been additional alignment costs if integration did not happen in the reference frame that was used to prepare and initiate actions (e.g., transforming the integrated representation into the reference frames used for action; see the **General discussion** section). Results of **Experiment 1**, however, strongly suggest that the reference frame during the first presentation was aligned with that during the second presentation in which the actual integration took place. In that case, no further alignment costs were expected in order to transform the integrated representation into the reference frame used to prepare and initiate actions, since these coincide anyway.

The results of [Experiment 1](#) suggest that the first representation was transformed to align it with the second representation in the different viewpoint condition. Consequently, participants seemed to integrate spatial information within the reference frame of the second presentation. This result is at odds with results from Kelly et al. (2010), in which participants used an already-established reference frame for encoding novel information. We will discuss potential reasons for this difference in detail in the [General discussion](#) section.

Integration in the study of Greenauer and Waller (2010) did not necessarily happen within a reference frame aligned with one of the intrinsic orientations of the object arrays, but was determined by the learning view and/or the overall pattern of the combined object array. In the present experiment, the tile patterns did not have clear intrinsic orientations; in fact, presentation patterns with three tiles aligned in a row were explicitly excluded. In addition, no general advantage for being aligned with an orientation relative to the room or to the carpet was observed in [Experiment 1](#). Consequently, the orientation of memory representations was likely to be determined by the experienced view. In light of lacking any clear intrinsic reference frame, integrating within a reference frame determined by experienced views is consistent with the results obtained by Greenauer and Waller (2010).

The results of [Experiment 1](#) showed, first, that spatial integration takes place within a single reference frame, even if spatial information was perceived from multiple viewpoints. Second, the results suggest that costs for aligning misaligned reference frames are likely to contribute to integration costs observed in various other experiments (Golledge et al., 1993; Hanley & Levine, 1983; Montello & Pick, 1993; Yamamoto & Shelton, 2008). Third, participants integrated within the reference frame of the second presentation. In [Experiment 2](#), we further examined the reason for selecting this reference frame.

Experiment 2

The results of [Experiment 1](#) indicated that participants encoded spatial information in orientation-dependent representations that were then aligned for integration. More specifically, the first representation was transformed to align with the second representation, and not vice versa. Two explanations for this specific transformation appear to be possible: (a) The *encoding hypothesis* states that spatial information is integrated with existing information when novel information is perceived and encoded. This implies that integration occurs in the reference frame from which novel information is perceived. (b) The *utilization hypothesis* states that integration happens when the task at hand

demand the integrated information. This implies that integration takes place in the reference frame that is used to prepare and initiate actions. [Experiment 2](#) was designed to test these hypotheses by separating the location from which novel information was encoded from the location from which this information was utilized. Participants learned one set of target locations from one location and a second set from a different location. In one condition, they moved back to the location of the first presentation, thus separating the locations of novel information encoding and utilization.

Method

Participants The task required a fair amount of training. In order to reduce this training, the same participants who participated in [Experiment 1](#) also participated in [Experiment 2](#).

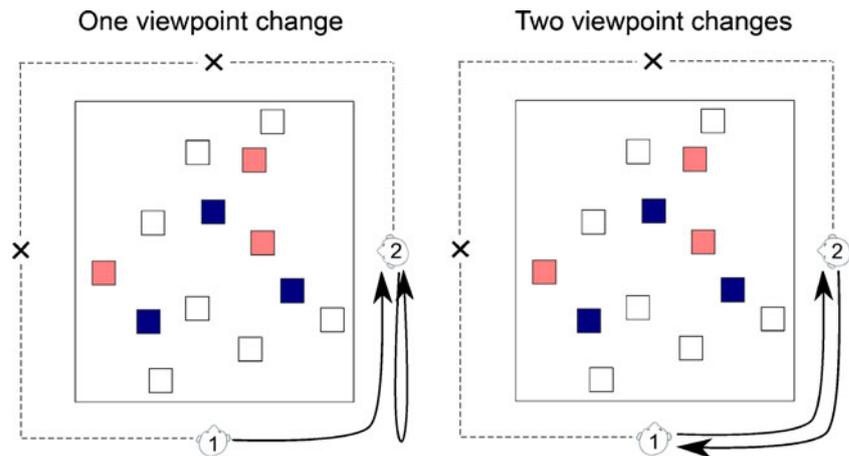
Procedure The experimental setup was identical to that of [Experiment 1](#). Each trial in [Experiment 2](#) started in the same way as a trial in the different viewpoint condition in [Experiment 1](#): Participants were presented three target tiles from one viewpoint; then, they moved to a second viewpoint at which they were presented with three additional target tiles. However, in contrast with [Experiment 1](#), participants were then required to move again: In the one-viewpoint-change condition, they walked to the next corner on their left-hand side and then came back to the location from which they observed the second presentation (see Fig. 5). In the two-viewpoint-changes condition, participants walked back to the first viewpoint. Thus, the starting location for the actual movement was either the location from which the first presentation was observed (two-viewpoint-changes condition) or from which the second presentation was observed (one-viewpoint-change condition). As in [Experiment 1](#), participants were asked to plan and navigate the shortest possible path to step on all six target tiles.

Twenty-four new patterns (subsets of 2x3 target tiles)—different from those in [Experiment 1](#)—were generated. All other methods, such as randomizing the order of patterns, balancing the two conditions, varying the start locations with respect to global orientation, and so on, were identical to those in [Experiment 1](#).

Predictions The different procedures in the one-viewpoint-change and the two-viewpoint-changes conditions allowed distinguishing between the encoding hypothesis and the utilization hypothesis. The crucial difference between the conditions is that the viewpoint of the second presentation and the startpoint coincide in the one-viewpoint-change condition, whereas they differ in the two-viewpoint-changes condition.

According to the encoding hypothesis, new spatial information is integrated directly when it is encoded. Therefore, integration happens in the reference frame

Fig. 5 Exemplary trial for **Experiment 2**: In both conditions, participants saw the first three tiles (dark blue squares) from “1” and the second three tiles (bright red squares) from “2.” After watching the second target tiles, they either walked to the corner and back (left side—one-viewpoint-change condition), or they returned to location “1” from where they saw the first three tiles (right side—two viewpoint changes). Then, they should have planned and walked the shortest route across these six tiles



coinciding with the perspective from which the novel information is perceived and encoded. In both conditions, participants have to transform the first presentation to align with the second presentation in order to integrate it when novel information is perceived. With respect to memory errors, we therefore do not expect an interaction between the order of presentation (tiles presented first and second) and the experimental condition.

The utilization hypothesis, however, makes different predictions for the two conditions: If integration happened in the reference frame from which an action is prepared and initiated, the first presentation has to be transformed to align it with the second presentation only in the one-viewpoint-change condition, since the start position is the same as the position from which the second presentation is learned. For the two-viewpoint-changes condition, in contrast, the second presentation has to be transformed into the reference frame of the first presentation from which participants initiate the action. Thus, the utilization hypothesis predicts an interaction between presentation order (tiles presented first vs. second) and the experimental conditions.

In addition to these hypotheses-specific interactions, main effects of both condition and presentation were expected. These effects, however, do not allow distinguishing between the encoding and utilization hypothesis, since both hypotheses predict these effects: First, presentation order is always confounded with general memory decay; thus, a main effect for presentation is expected. Both hypotheses predict higher error rates in the two-viewpoint-changes condition than in the one-viewpoint-change condition: The encoding hypothesis states that an additional transformation is required in the two-viewpoint-changes condition, but not the one-viewpoint-change condition. This is because the integrated representation has to be transformed from the second viewpoint back to the first viewpoint from where actions are initiated in the two-viewpoints-changes condition only. The utilization hypothesis, on the other hand, states that only one representation has to be transformed into the orientation of

acting (either the first or the second) in both conditions. In the two-viewpoint-changes condition, however, this transformation has to be executed while keeping the other representation in memory, which should yield a higher error rate as compared with the one-viewpoint-change condition.

Results and discussion

As in **Experiment 1**, participants showed good path planning performance: On average, the chosen paths were 7.2% longer than the optimal solution (see Fig. 3). This performance is clearly better than predicted if no integration happened (i.e., stepping on the tiles of the different presentations successively: $t(15) = 16.8$, $p < .001$). Again, this result suggests that participants effectively integrated spatial information perceived from the different viewpoints.

The main purpose for **Experiment 2** was to examine whether integration happens while encoding novel information (encoding hypothesis) or when the integrated information is required (utilization hypothesis). As predicted by the utilization hypothesis, but not the encoding hypothesis, a significant interaction between presentation and condition was observed, $F(1, 740) = 5.08$, $p = .025$. Figure 6, in which error rates for the two conditions are plotted separately for the first and the second presentation, demonstrates this interaction: The error rate for tiles presented second in the two-viewpoint-change condition is higher than that of the one-viewpoint-change condition [difference $M = 0.31$, $SD = 0.84$; $F(1, 364) = 12.8$, $p < .001$]. No such difference could be established for tiles in the first presentation [difference $M = 0.05$, $SD = 0.78$; $F(1, 364) < 1$]. The significant interaction and the increase in error rate for the second presentation were predicted by the utilization hypothesis, but not by the encoding hypothesis. This result suggests that participants transformed the representation of tiles presented second to align it with the first representation in the two-viewpoints-change condition. This is most likely due to fact that participants were required to start navigating

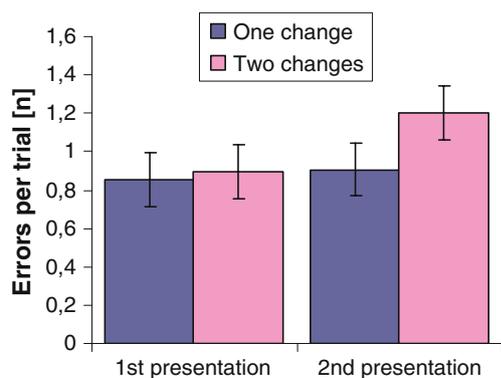


Fig. 6 Differences in error rate between conditions for tiles presented first and second

from the location from which they were presented with the first presentation.

Contrary to [Experiment 1](#), there was no difference between the first and second presentation in the one-viewpoint-change condition (see [Fig. 6](#)). Both hypotheses predicted transformations and thus higher errors for the tiles of the first as compared with the second presentation. A lack of power or of alternating strategies within this condition might be a reason for the lack of this effect. Still, within-condition comparisons across presentations are not very meaningful in the present setting, since they are confounded with memory decay, interference with walking, and so on. However, comparing error rates for the first presentations between conditions should have produced a difference between conditions: According to the utilization hypothesis, the first presentation is transformed to align it with the second presentation in the one-viewpoint-change condition. In the two-viewpoint-changes condition, in contrast, participants had to transform the second presentation toward the first presentation. Accordingly, we expected a difference between conditions not only for the second presentation (as observed), but also for the first presentation (in the opposite direction). This difference might have been occluded by an opposing effect. In the two-viewpoint-changes condition, the first presentation had to be kept in memory while the second presentation was transformed. This transformation might have interfered with memory retention, thus occluding differences due to transformation versus nontransformation. Note that the encoding hypothesis also predicted a difference between conditions in the first presentation, however, in the opposite direction. The results for the first presentation were predicted by neither of the two hypotheses. The utilization hypothesis, however, predicted both the interaction between condition and presentation and the observed difference between conditions for the second presentation.

Not surprisingly, we also found a main effect of condition, which was predicted by both hypotheses, $F(1,$

740) = 8.76, $p = .003$. Participants made more errors in the two-viewpoint-changes condition ($M = 0.82$, $SD = 0.77$ errors per presentation) than in the one-viewpoint-change condition ($M = 1.1$, $SD = 0.85$). In accordance with the utilization hypothesis, the data suggest that transforming the first presentation before perceiving the second presentation (one-viewpoint-change condition) is easier than transforming the second presentation while keeping the first presentation in memory at the same time (two-viewpoint-changes condition). This would also explain the main effect of presentation, $F(1, 740) = 10.3$, $p = .001$, which suggests a higher overall error rate for tiles of the second presentation ($M = 1.1$, $SD = 0.85$) than for tiles of the first presentation ($M = 0.85$, $SD = 0.78$). The direction of the main effect of presentation was opposite of that in [Experiment 1](#), in which a strong advantage for tiles presented second was found. In [Experiment 2](#), however, participants could not start navigating immediately after the second presentation: Rather, they first had to perform another movement. Therefore, participants might have relied on long-term memory for both views in [Experiment 2](#), whereas they relied on short-term memory for the second presentation in [Experiment 1](#), yielding lower error rates for these tiles. There was no main effect of condition for latency, $F(1, 267) < 1$.

There was a main effect of order, errors $F(1, 12) = 12.7$, $p = .004$, time, $F < 1$, indicating better memory performance when conducting the two-viewpoint-changes condition first. Since order was balanced between participants and included as a factor in the analysis, this effect does not provide an alternative explanation for the results discussed previously; also, we did not observe any significant interaction between order and condition or presentation. No gender difference was observed [errors, $F(1, 12) = 2.34$, $p = .152$; time, $F < 1$], and performance did not differ between the four different start locations (errors and time: $F < 1$).

General discussion

The integration of spatial information perceived at different times and from different viewpoints is a common task in spatial learning, navigation, and wayfinding. However, surprisingly little is known about the underlying processes and mechanisms. In the present experiments, we demonstrated that spatial representations encoded from different viewpoints have to be aligned with each other in order to be integrated. This alignment process results in transformation costs. The results of [Experiment 1](#) demonstrated higher memory costs (i.e., a higher error rate) when integrating spatial information acquired from two different viewpoints, as compared with integrating two representations acquired from the same viewpoint. Although several earlier experiments that examined the integration of spatial information

have described integration costs (Golledge et al., 1993; Hanley & Levine, 1983; Montello & Pick, 1993; Yamamoto & Shelton, 2008), our results suggest that part of these integration costs are the result of aligning differently oriented representations.⁴

The observed alignment costs suggest that spatial information is integrated within a single reference frame, which requires the reference frames of both representations to be brought together. This has been proposed theoretically (Greenauer & Waller, 2010; McNamara et al., 2008; Meilinger, 2008) and has been demonstrated empirically for spatial learning from a single viewpoint (Greenauer & Waller, 2010). In the present work, we extend these results toward more natural learning situations in which spatial information is experienced from different viewpoints.

The second aim of the present study was to investigate which reference frame was chosen for integration. In [Experiment 1](#), an increase in error rate for target tiles of the first presentation was observed only in the different viewpoint condition, in which the two presentations were perceived from different locations. This result suggests that the representation of the first presentation was transformed toward the reference frame of the second, from which participants perceived novel information and started acting.

In [Experiment 2](#), participants integrated in the reference frame from which they used spatial information—that is, from which they started to act—rather than in the reference frame in which novel information was presented. In the crucial two-viewpoint-changes condition, they primarily transformed novel information from the second presentation toward the reference frame of the first, from which they started acting. Thus, spatial information perceived at different times and from different locations seems to be integrated when required, rather than when novel information is perceived. This is consistent with the literature demonstrating that participants integrate spatial information immediately on perception only when knowing this will help them and when provided with sufficient time to do so. Otherwise, participants kept representations separate (Golledge et al., 1993; Hanley & Levine, 1983; Holding & Holding, 1989; Moar & Carleton, 1982; Maguire et al., 1996; Montello & Pick, 1993; Yamamoto & Shelton, 2008).

Aligning spatial information in the orientation in which it is to be utilized might be part of a more general strategy of minimizing alignment costs. Keeping misaligned spatial representations separate removes the need for unnecessary transformations. This is particularly true when considering

alignment for action in addition to alignment for integration. Alignment for action occurs, for example, when a memory representation of an environment is misaligned with a navigator's current orientation within this environment. Judgments of relative directions paradigms use such misalignments to estimate the orientation of a memory representation; imagined orientations aligned with the representation yield better performance than do misaligned representations (Greenauer & Waller, 2010; Kelly & McNamara, 2008; Kelly et al., 2010; Mou & McNamara, 2002; Shelton & McNamara, 2001). Integrating spatial information only within the reference frame required to act erases such additional transformation costs, since the integrated spatial information is already aligned with the reference frame used for action. Consequently, keeping representations separate and aligning them in the orientation of usage only when integration is required appears to be an efficient strategy for minimizing overall alignment costs.

However, Kelly et al. (2010) demonstrated encoding of novel information within a reference frame that was already established. It thus seems that the reference frame used for integration or for the encoding of novel information depends on the exact circumstances. Particularly, presentation time as well as test orientations may be relevant for differences in reference frame selection. In the experiments from Kelly and colleagues, encoding time was in the range of minutes. Participants were thus able to choose the reference frame most appropriate to encode novel spatial information—in this case, the existing reference frame—and to transform novel information into this reference frame without time pressure and probably also without observable costs (Hanley & Levine, 1983; Yamamoto & Shelton, 2008). During testing, they then retrieved information from long-term memory. In the present experiments, in contrast, encoding time was limited and may have been too short to encode information in a reference frame different from the one provided by the current view. Also, testing immediately followed the second presentation in [Experiment 1](#), which makes usage of a short-term representation for the corresponding target tiles likely. The involvement of short-term memory, however, is unlikely in the study of Kelly and colleagues. In addition, participants in the present experiments were aware of the test orientation, which was also familiar, whereas participants in the experiments by Kelly and colleagues were tested from multiple novel and unexpected orientations. Given the limited encoding time and their knowledge about future test orientations in the present study, participants could have transformed their memories of the target locations into the reference frame used for action already during walking—for example, by updating a working memory representation of the target locations (spatial updating). For future research, it would be an interesting challenge to systematically address how

⁴ In the experiment of Yamamoto and Shelton (2008), object layouts were perceived from one viewpoint only. Still, alignment costs could have occurred if these layouts differed in their intrinsic orientation (cf. Greenauer & Waller, 2010). Unfortunately, this information was not provided in the article.

encoding time, knowledge about future test orientations, the use of different memory systems, and the possibility of updating locations influence the selection of reference frames for integration. Together, results from the present and earlier studies suggest a rather flexible mechanism for the integration and encoding of novel spatial information.

Our results suggest that spatial information was transformed in order to be integrated. Such transformations in working memory can be achieved either by mentally rotating a representation while standing at one location (e.g., Schmidt et al., 2007) or by continuously updating the representation while moving to a different viewpoint (e.g., Wang, 2004). From the present data, we cannot distinguish whether participants mentally rotated their representations, updated their representations, or did both. Note, however, that obligatory or automatic updating is not consistent with our data (for related critic on automatic updating, see Amonrim, Glasauer, Corprinet, & Berthoz, 1997; Wang, 2004). The results of Experiment 1, for example, can be explained only by assuming that participants ignored movements between target presentations in the same viewpoint condition (i.e., they did not automatically update the locations) and instead relied on memory encoded before movement. If participants automatically updated all locations, no differences would have been expected between conditions. The combination of updating mechanisms and view-dependent memory, on the other hand, are consistent with the present data. The main aim of the present experiments, however, was to demonstrate, first, that aligning was necessary for integration at all, and second, to specify the reference frame within which this might occur.

Many earlier experiments demonstrated that the learning of spatial layouts can lead to the establishment of a stable reference frame relative to the surrounding room or the object array (Kelly & McNamara, 2008; Mou & McNamara, 2002; Shelton & McNamara, 2001; Valiquette & McNamara, 2007). To control for such effects, we systematically varied the start locations relative to the room and the carpet in both experiments. We did not observe any systematic differences between the different start orientations. One reason for this may be that the learning times in our experiments were much shorter, in the range of seconds, as compared with the aforementioned studies in which participants learned the layouts for several minutes. Furthermore, we did not test for oblique misalignments relative to the room/carpet (i.e., 45°, 135°), which often influence performance more than orthogonal misalignments (90°, 180°) in these experiments.

Conclusions

Despite the fact that the integration of spatial information across different viewpoints is a frequent spatial task in

everyday navigation, very little is known about it. In the present study, we provided first results concerning the process of integration. Taken together, the results suggest, first, that navigators encode multiple representations that have to be aligned for integration. Second, the results suggest that the integration happens within a single reference frame, and third, that the integration—at least in the present circumstances—is conducted when needed for action rather than during encoding, which might serve to circumvent alignments.

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 P. Leboucher, M. Ehrette, J. Campos, J. Butler, G. Griffin, A. Henson, and the participants for their help.

References

- Amonrim, M.-A., Glasauer, S., Corprinet, K., & Berthoz, A. (1997). Updating an object's orientation and location during nonvisual navigation: A comparison between two processing modes. *Perception & Psychophysics*, *59*, 404–418.
- 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.
- Brockmole, J. R., Wang, R. F., & Irwin, D. E. (2002). Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 315–334.
- Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*, 340–375.
- Christou, C. G., & Bühlhoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, *27*, 996–1007.
- Golledge, R. G., Ruggles, A. J., Pellegrino, J. W., & Gale, N. D. (1993). Integrating route knowledge in an unfamiliar neighbourhood: Along and across route experiments. *Journal of Environmental Psychology*, *13*, 293–307.
- Greenauer, N., & Waller, D. (2010). Micro- and macro-reference frames: Specifying the relations between spatial categories in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 938–957.
- Hanley, G. L., & Levine, M. (1983). Spatial problem solving: The integration of independently learned cognitive maps. *Memory & Cognition*, *11*, 415–422.
- Holding, C. S., & Holding, D. H. (1989). Acquisition of route network knowledge by males and females. *The Journal of General Psychology*, *116*, 29–41.
- Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, *52*, 93–129.
- Jiang, Y. (2004). Time window from visual images to visual short-term memory: Consolidation or integration? *Experimental Psychology*, *51*, 45–51.
- Kelly, J. K., Avraamides, M., & McNamara, T. P. (2010). Reference frames influence spatial memory development within and across sensory modalities. In C. Hölscher, T. F. Shipley, M. Olivetti

- Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), *Spatial cognition VII* (pp. 222–233). Berlin: Springer.
- Kelly, J. K., & McNamara, T. P. (2008). Spatial memories of virtual environments: How egocentric experience, intrinsic structure, and extrinsic structure interact. *Psychonomic Bulletin & Review*, *15*, 322–327.
- Maguire, E. A., Burke, T., Phillips, J., & Staunton, H. (1996). Topographical disorientation following unilateral temporal lobe lesions in humans. *Neuropsychologia*, *34*, 993–1001.
- McNamara, T. P., Slucenski, J., & Rump, B. (2008). Human spatial memory and navigation. In J. H. Byrne (Ed.), *Learning and Memory: A comprehensive reference, Vol. 2. Cognitive psychology of memory* (pp. 157–178). Oxford: Elsevier.
- Meilinger, T. (2008). The network of reference frames theory: A synthesis of graphs and cognitive maps. In C. Freksa, N. S. Newcombe, P. Gärdenfors, & S. Wöflfl (Eds.), *Spatial Cognition VI* (pp. 344–360). Berlin: Springer.
- Moar, I., & Carleton, L. R. (1982). Memory for routes. *The Quarterly Journal of Experimental Psychology*, *34A*, 381–394.
- Montello, D. R., & Pick, H. L. (1993). Integrating knowledge of vertically aligned large-scale spaces. *Environment and Behavior*, *25*, 457–484.
- Mou, W., & McNamara, T. M. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*, 162–170.
- Presson, C. C., DeLange, N., & Hazelrigg, M. D. (1987). Orientationspecificity in kinesthetic spatial learning: The role of multiple orientations. *Memory & Cognition*, *15*, 225–229.
- Schmidt, D., et al. (2007). Visuospatial working memory and changes of point of view in 3D space. *Neuroimage*, *36*, 955–968.
- Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology*, *43*, 274–310.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. Thousand Oaks: Sage.
- Valiquette, C., & McNamara, T. P. (2007). Different mental representations for place recognition and goal localization. *Psychonomic Bulletin & Review*, *14*, 676–680.
- Wang, R. F. (2004). Between reality and imagination: When is spatial updating automatic? *Perception & Psychophysics*, *66*, 68–67.
- Wiener, J. M., Ehbauer, N., & Mallot, H. A. (2009). Planning paths to multiple targets: Memory involvement and planning heuristics in spatial problem solving. *Psychological Research*, *77*, 644–658.
- Wiener, J. M., Lafon, M., & Berthoz, A. (2008). Path planning under spatial uncertainty. *Memory & Cognition*, *36*, 495–504.
- Yamamoto, N., & Shelton, A. L. (2008). Integrating object locations in the memory representation of a spatial layout. *Visual Cognition*, *16*, 140–143.