ARTICLE IN PRESS

Available online at www.sciencedirect.com
ScienceDirect

COGNITION

ELSEVIER

Cognition xxx (2007) xxx–xxx

www.elsevier.com/locate/COGNIT

Intrinsic frames of reference and egocentric viewpoints in scene recognition

Weimin Mou [a,*], Yanli Fan [a], Timothy P. McNamara [b], Charles B. Owen [c]

[a] *State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, 10A Datun Road, Beijing 100101, China*
[b] *Department of Psychology, Vanderbilt University, United States*
[c] *Department of Computer Science and Engineering, Michigan State University, United States*

Received 31 July 2006; revised 14 April 2007; accepted 16 April 2007

Abstract

Three experiments investigated the roles of intrinsic directions of a scene and observer's viewing direction in recognizing the scene. Participants learned the locations of seven objects along an intrinsic direction that was different from their viewing direction and then recognized spatial arrangements of three or six of these objects from different viewpoints. The results showed that triplets with two objects along the intrinsic direction (intrinsic triplets) were easier to recognize than triplets with two objects along the study viewing direction (non-intrinsic triplets), even when the intrinsic triplets were presented at a novel test viewpoint and the non-intrinsic triplets were presented at the familiar test viewpoint. The results also showed that configurations with the same three or six objects were easier to recognize at the familiar test viewpoint than other viewpoints. These results support and develop the model of spatial memory and navigation proposed by Mou, McNamara, Valiquette, and Rump [Mou, W., McNamara, T. P., Valiquiette C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 142–157].

* Corresponding author. Tel.: +86 10 6483 7040; fax: +86 10 6487 2070.
E-mail address: mouw@psych.ac.cn (W. Mou).

0010-0277/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2007.04.009

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

Keywords: Scene recognition; Intrinsic frame of reference; Egocentric viewpoint; Spatial memory; Visual memory

1. Introduction

Successful human navigation relies on mental representations of spatial relations among important elements in the surrounding environment. There are at least two primary spatial processes calling on the representation of spatial relations. One such process is to judge the direction and distance of landmarks and navigational goals, which may be in the immediate environment or unseen. A second key process is to recognize visually the spatial relations in the immediate environment to maintain one's orientation or to reorient. The goal of this project was to investigate whether a common spatial representation or two different spatial representations supported these two spatial behaviors.

A decade ago, McNamara and his colleagues (Diwadkar & McNamara, 1997; Shelton & McNamara, 1997) obtained evidence indicating that both spatial behaviors used the same egocentric spatial representations. In the experiments of Diwadkar and McNamara, participants learned the locations of 7 objects on a table from a single viewpoint. Participants then made old-new recognition judgments on the test scenes consisting of pictures of the learned configuration taken at the familiar viewpoints and several novel viewpoints and pictures of novel configurations of the same objects. For the pictures of the learned configuration, recognition performance was better for the experienced views than for the novel views. In parallel, Shelton and McNamara reported that judgments of relative direction were learning orientation dependent. Participants learned the locations of several objects on the floor from two orthogonal viewpoints successively. Participants then moved to a different room and made judgments of relative directions ("Imagine you are standing at X, facing Y, please point to Z") using spatial memory. The results showed that judgments of relative direction were better at the imagined headings parallel to the learning directions than at the novel imagined headings.

In the past decade, many studies have replicated the finding that human visual scene recognition is viewpoint dependent, such that recognition performance is better for experienced views than for novel views (e.g., Burgess, Spiers, & Paleologou, 2004; Christou & Bülthoff, 1999; Diwadkar & McNamara, 1997; Shelton & McNamara, 2001, 2004a, 2004b; Simons & Wang, 1998; Wang & Simons, 1999). However, recent evidence suggests that judgments of relative direction may not be learning viewpoint dependent; instead they may be intrinsic-orientation dependent, such that pointing performance is better for novel imagined headings parallel to directions highlighted by the experimenter (Mou & McNamara, 2002).[1]

[1] Intrinsic-orientation dependent pointing judgments suggest that the spatial representations in long-term memory that support the pointing judgments are allocentric. Clearly the allocentric spatial representation must be translated into egocentric coordinates by aligning the egocentric front with the imagined facing direction when participants point to the target object.

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

In one experiment of Mou and McNamara's (2002) study, participants viewed an array of objects from a single viewpoint (315°) but were instructed to learn the layout along an intrinsic direction (0°), which was 45° different from the viewing direction. After learning, participants moved to a different room and made judgments of relative direction using their memories of the spatial layout of the objects. Participants were better able to perform the task from a novel imagined heading (0°), which was parallel to the intrinsic direction instructed, than from the imagined heading they actually experienced (315°). Mou and McNamara proposed that people use intrinsic frames of reference to specify locations of objects in memory. More specifically, the spatial reference directions, which are established to represent locations of objects, are not egocentric (e.g., Shelton & McNamara, 2001). Instead the spatial reference directions are intrinsic to the layout of the objects. There is an infinite number of possible intrinsic directions inside a layout of several objects. A small number of them (1 or 2 typically) is selected using cues available to the participant, such as the participant's viewing perspective, properties of the layout (e.g., the symmetric axis of a layout), the structure of the environment (e.g., geographical slant), and even instructions.

Dissociations in the patterns of results in scene recognition and judgments of relative direction have also been demonstrated (e.g., Shelton & McNamara, 2001, 2004a, 2004b; Valiquette & McNamara, in press). For example, in one experiment, Shelton and McNamara (2004b) had one participant (the director) view a display of objects from a single perspective and describe the display to a second participant (the matcher) from a perspective that differed from the viewing perspective. The director's memory for the spatial layout was tested using judgments of relative direction and scene recognition. The results showed that performance for judgments of relative direction was best at the imagined heading parallel to the described view, whereas the performance for scene recognition was best at the visually perceived view.

One hypothesis that has been advanced to explain this dissociation is that two independent spatial representations may be formed when participants learn a spatial layout visually (e.g., Valiquette & McNamara, in press). One of these representations seems to preserve interobject spatial relations, and is used to make judgments of relative direction, whereas the other is a visual memory of the layout, and supports scene recognition. Spatial memory is organized with respect to intrinsic frames of reference as suggested by Mou and McNamara (2002), so judgments of relative direction appear to be intrinsic-orientation dependent. Visual memory is formed from the experienced viewpoints, so scene recognition is viewpoint dependent.

In this project, we propose and test an alternative hypothesis to reconcile the contrasting results from visual scene recognition and judgments of relative direction. This hypothesis derives from and develops the model of spatial memory and navigation that was proposed by Mou, McNamara, Valiquette, and Rump (2004; see also, Waller & Hodgson, 2006). According to this model, the human navigation and spatial representation system comprises two subsystems: The egocentric subsystem computes and represents transient self-to-object spatial relations needed to control locomotion (e.g., walking through apertures, such as doorways). These spatial relations are represented at sensory-perceptual levels and decay relatively rapidly in the

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

absence of perceptual support or deliberate rehearsal. The environmental subsystem is responsible for representing the enduring features of familiar environments. In this subsystem interobject spatial relations are represented in terms of a small number (typically 1 or 2) of intrinsic reference directions or axes (e.g., Mou & McNamara, 2002). Furthermore the location and orientation (viewing direction) of the observer are also originally represented with respect to the same intrinsic reference direction. When people move in the environment they update their location and orientation with respect to the intrinsic frame of reference.

This theoretical framework is illustrated in Fig. 1a. Participants' study viewing direction is illustrated by the solid arrow and the intrinsic reference direction is illustrated by the dashed arrow. The angular directions from object 1 to object 3 (α_{13}), from object 1 to object 2 (α_{12}), and from object 1 to object 4 (α_{14}) are all specified with respect to the intrinsic reference direction. Judgments of angular directions with respect to a direction parallel to the intrinsic reference direction are easier than judgments of angular directions with respect to other directions because the angular directions with respect to the intrinsic reference direction are represented but the angular directions with respect to other directions are not represented and need to be inferred (e.g., Klatzky, 1998). These inferential processes introduce observable costs in terms of latency and error. For example, people are better assessing the angular direction from object 1 to object 3 with respect to the direction from object 1 to object 2 than with respect to the direction from object 1 to object 4. This model can readily explain the intrinsic-orientation dependent performance in judgments of relative direction.

We hypothesized that scene recognition uses the same spatial representation as judgments of relative direction. The goal in scene recognition is to determine whether the layout of objects in the test scene is the same as the represented layout of the objects in the remembered study array. The processes involved in this decision will be facilitated if the test scene can be represented in a manner that is congruent with

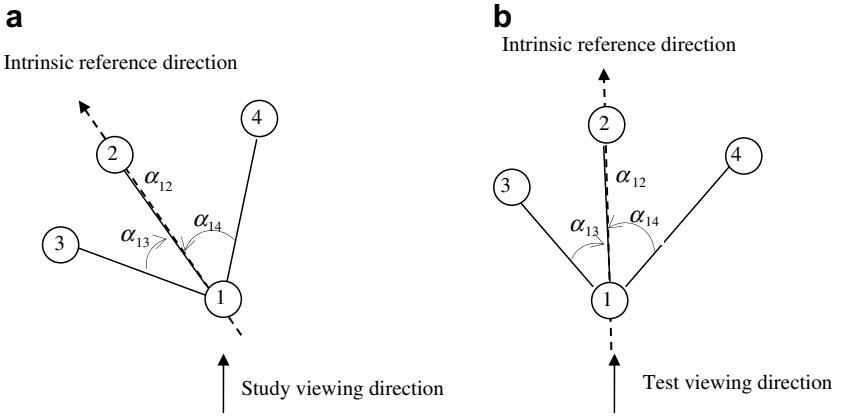

Fig. 1. The model of spatial memory for scene recognition

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

the mental representation of the study array. Our conjecture is that the reference direction in the mental representation can be identified more efficiently in a test scene that contains at least two objects on an axis parallel to the reference direction than in a test scene that does not contain any objects parallel to the reference direction. Successful identification in the test scene of the represented reference direction in turn facilitates the decision about whether it matches the study array. Hence the above model makes a novel prediction about performance in scene recognition: People will be better able to recognize a scene containing at least two objects that lie on an axis parallel to the intrinsic reference direction than a scene that does not have two objects falling on a line parallel to the intrinsic reference direction. For example, recognizing the triplet of objects 1, 2, 3 in Fig. 1a would be easier than recognizing the triplet of objects 1, 3, 4 because in the former triplet objects 1 and 2 lie on the intrinsic reference direction, whereas in the latter triplet there are not two objects falling on a line parallel to the intrinsic reference direction.

According to the model proposed by Mou et al. (2004), people also represent their viewing direction at study with respect to the intrinsic reference direction and update their orientation with respect to the intrinsic reference direction when they move around the studied layout. If people are tested at the study position and the test view is different from the study view, the reference direction of the target scene is oriented differently from the reference direction in the mental representation. We assume that when the reference direction of the target scene is oriented differently from the represented reference direction in memory (i.e., the test view is different from the study view), people need to align these two intrinsic directions (e.g., Ullman, 1996). For example, participants need to align the intrinsic reference direction defined by objects 1 and 2 at test (Fig. 1b) with the intrinsic reference direction defined by objects 1 and 2 at study (Fig. 1a). We also assume that processing costs are introduced during the alignment of the intrinsic reference directions. Because participants do not need to align the intrinsic reference direction when the test view is the same as the study view, scene recognition appears to be viewpoint dependent when the intrinsic orientation effects are constant across different testing viewpoints.

This prediction also holds if people are disoriented between study and test or are tested in a room different from the study room. Under such conditions, people lose track of their orientation with respect to the studied layout and retrieve the represented intrinsic reference direction as if they were still facing in the study viewing direction (for evidence, see Mou, McNamara, Rump, & Xiao, 2006). Hence the spatial relations between their body orientation and the represented intrinsic reference direction do not change from study to test. In other words, the represented reference direction with respect to the test viewing direction is the same as the represented reference direction with respect to the study viewing direction.

Why was intrinsic-orientation dependence in scene recognition not observed previously (e.g., Shelton & McNamara, 2004b; Valiquette & McNamara, in press)? Shelton and McNamara have shown that scene recognition is viewpoint dependent even when the intrinsic directions were dissociated from the study viewing direction. In previous studies, however, the same configuration was displayed from different viewpoints. We hypothesized that the intrinsic-orientation effect would disappear

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

if the same configuration was tested from different viewpoints because the interobject spatial relations with respect to the intrinsic reference direction would be the same across all viewpoints. Instead a viewpoint dependent pattern should appear because the intrinsic reference direction in the mental representation and the identified intrinsic reference direction in the target scene would need to be aligned when the study view was different from the test view.

In summary, we hypothesized that the inconsistent findings from the paradigms of scene recognition and judgments of relative direction might not be due to the different tasks that rely on different representations. Rather we hypothesized that scene recognition, just as judgments of relative direction, also relies on an allocentric spatial memory that preserves interobject spatial relations and the study viewing direction of the observer with respect to the intrinsic frame of reference. Scene recognition is intrinsic-orientation dependent in the sense that intrinsic triplets, in which two objects lie along the intrinsic reference direction, are easier to recognize than non-intrinsic triplets, in which there were no two objects falling in the intrinsic reference direction. Scene recognition is viewpoint dependent in the sense that the same configurations are easier to recognize when the test view is the same as the study view than when they are different.

Three experiments were conducted to test these hypotheses. Experiments 1A and 1B investigated whether scene recognition could also be intrinsic-orientation dependent as observed in judgments of relative direction when intrinsic triplets were tested at the viewpoint parallel to the intrinsic direction, whereas non-intrinsic triplets were tested at the viewpoint parallel to the study viewing direction. Experiment 2 tested whether the intrinsic-orientation effect would disappear but the learning viewpoint effect would appear when the same configurations were tested from different viewpoints. Experiment 3 was designed to factorially dissociate the intrinsic-orientation effect and learning viewpoint effect in scene recognition.

2. Experiment 1A

In Experiment 1A, participants learned locations of objects from the point of view labeled 315° (Fig. 2). They were asked to learn the locations of the objects according to the columns in the 0–180° direction, as indicated by the experimenter (e.g., clip–hat; glue–wood–ball; lock–candle) and were required to name and point to the objects in a manner consistent with this organization. Hence the 0° to −180° axis was established as the intrinsic orientation. The same sets of triplets used by Mou and McNamara (2002) were presented visually at the viewpoint specified by the imagined heading in judgments of relative direction. For example, corresponding to the trial ''Imagine you are standing at the wood, facing the ball, please point to the lock'', the triplet wood–ball–lock was visually presented at the viewpoint established by the wood and the ball (0° in Fig. 2). The main purpose of the experiment was to determine whether scene recognition relies on visual memory that is viewpoint dependent or spatial memory that is intrinsic-orientation dependent, in particular, whether performance would be best for the view of 315° or 0°.

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

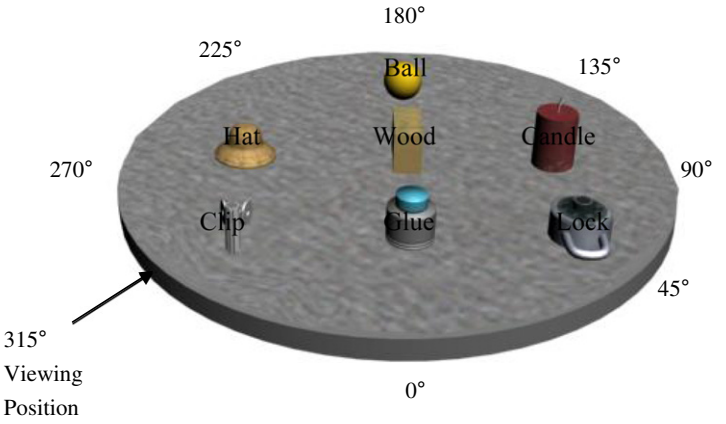

Fig. 2. The layout of objects used in Experiments 1A, 2, and 3. (315° indicates the study viewing position; 0° indicates the intrinsic direction.)

2.1. Method

2.1.1. Participants

Twenty university students (10 men, 10 women) participated for the return of monetary compensation.

2.1.2. Materials and design

A cylindrical room (3 m in diameter) with walls covered in black curtains was used as the learning room. The layout consisted of a configuration of seven common objects whose longest dimension was approximately 5 cm (see Fig. 2). The objects were placed on a circular table covered by a grey mat (50 cm in diameter, 48 cm above the floor). The configuration was the same as that used by Mou and McNamara (2002) except that the size was smaller. The distance between the clip and the hat was 14.14 cm. The table was placed in the middle of the learning room. There was a chair (seated 42 cm high) located at the viewing position, with the back of the chair 100 cm from the middle of the table.

Another room on the same floor was used as the testing room, where a table and a chair identical to those in the learning room were placed. The distance between the chair and the table was identical to that in the learning room. Virtual objects instead of real objects were displayed on a virtual table that exactly occupied the location of the real table with a fiducial-based video see-through virtual reality system (Owen, Tang, & Xiao, 2003). All of the virtual objects and the virtual table were virtual analogs of the real ones in the learning room and were presented with the exact scale. The virtual reality system consists of a light (about 7 oz) glasses-like I-glasses PC/SVGA Pro 3D head-mounted display (HMD, I-O Display Systems, Inc. California) with a small video camera attached, and a group of 4 fiducials printed on a paper on the top of the table. The HMD supplied identical images to both eyes at a resolution

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

of 800 by 600 pixels and a field of view (FOV) of 26° diagonally for each eye. The virtual objects and the virtual table were rendered with an ATI Radeon graphics accelerator that updated the graphics on the display at 60 Hz. The virtual objects and the virtual table were presented on the origin of the coordinates (superimposed at the center of the table) which was defined by the groups of fiducials and could be recognized by the video camera mounted on the HMD. The participants were required to look at the center of table, so the virtual objects and the virtual table could be seen at the center of the FOV through the HMD.

The recognition test materials consisted of 48 target scenes and 48 distractor scenes. Each target scene consisted of a view of the array of objects with only three of the seven objects present. Six sets of three objects were used for each of eight views (0–315° in 45° increments; see Fig. 2). For example, the trials for the view of 0° (the instructed intrinsic direction) were: wood–ball–clip, wood–ball–lock, lock–candle–wood, lock–candle–hat, clip–hat–wood, and clip–hat–glue. The test trials at 315° (the study viewing direction) were: hat–ball–clip, hat–ball–lock, glue–candle–clip, glue–candle–hat, clip–wood–glue, clip–wood–candle. The three locations in each target scene corresponded to locations used by Mou and McNamara (2002) in judgments of relative direction. For example, the 0° target scene containing wood, ball, and lock corresponded to a judgment of relative direction in Mou and McNamara's study of the following form: "Imagine you are standing at the wood, facing the ball. Point to the lock." Hence, two objects in each target scene defined an axis parallel to the point of view represented by the scene. The distractor scenes were created from the target scenes by mirror reflecting the target scenes about the viewing axis at test.[2]

The primary independent variable was the test view (0–315° in 45° increments; see Fig. 2). The dependent measures were response latency and accuracy. Response latency was measured as the time from presentation of the test configuration to the target response.

2.1.3. Procedure

2.1.3.1. Learning phase. Before entering the study room, each participant was instructed to learn the locations of the objects for a scene recognition test and given one configuration of four objects as a practice so that the participant would be familiar with the procedure. The participant was blindfolded and led to be seated in the chair at the viewing position (315° in Fig. 1) in the learning room. The blindfold was removed and the participant was asked to learn the locations of the objects according to the columns in the 0–180° direction, as indicated by the experimenter (e.g., clip–hat; glue–wood–ball; lock–candle). The participant viewed the display for 30 s

[2] We used mirror-image distractors because they are uniquely associated with target scenes, unlike random configurations of objects or distractors constructed by switching the positions of objects. Participants were never informed that the distractors were mirror-images and they were asked to decide whether the three objects in the test scene were in the correct spatial configuration, regardless of view point.

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

before being asked to name and point to the objects with eyes closed. All participants named and pointed to the objects in an order consistent with the intrinsic axis. The order in which columns of objects were identified was changed from trial to trial. Five study-test trials were used.

2.1.3.2. Testing phase. After learning the spatial layout, participants were taken to the testing room. They were seated in the testing chair and faced the testing table on which the virtual objects and the virtual table would be presented. Participants wore the head mounted display and held a mouse, which was mounted firmly on a bar stool placed to the right of the test chair, with their right hand. Participants were required to look at the center of the real table through the HMD. Then the camera image of the real environment was turned off. Each test scene was presented on the virtual table, which occupied the location of the real table, once the experimenter pressed the space key on a keyboard. The scene disappeared once the participants pressed the mouse buttons (right button for target configurations and left button for distractor configurations). Participants were asked to decide whether the three objects in the test scene were in the correct spatial configuration, regardless of viewpoint. Participants were instructed to make their responses as rapidly as possible without sacrificing accuracy.

2.2. Results

Only the responses to the target configurations were analyzed. Response latency of the correct responses was analyzed in an ANOVA with one term for test view (0–315° in 45° steps). In this and all subsequent experiments, error rate was low on every test view (less than one error at each viewpoint) and showed the same general pattern as response latency. There was no evidence of speed–accuracy trade-offs. In the interest of brevity, we only report response latency.

Mean response latency is plotted in Fig. 3 as a function of test view. As shown in the figure, the major finding was that participants were quicker recognizing the target

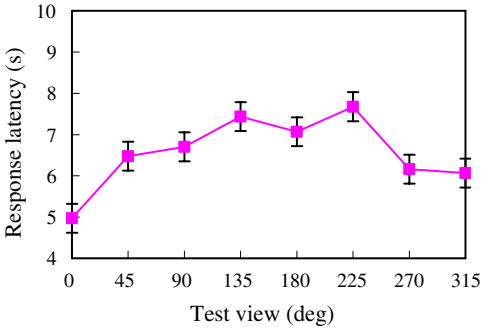

Fig. 3. Response latency in scene recognition as a function of test view in Experiment 1A. (All participants viewed the layout from 315° and were instructed to learn it along the 0–180° axis. Error bars are confidence intervals corresponding to ±1 standard error of the mean as estimated from the ANOVA.)

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

scenes with the view of 0°, which corresponded to the intrinsic axis, than with the view of 315°, which corresponded to the study view. In other words, performance in scene recognition was better for a novel view than for a familiar view.

This conclusion was supported by statistical analyses. The overall effect of test view was significant, $F(7,133) = 6.06$, $p < .001$, $MSE = 2.46$. Pairwise comparisons showed that response latency was shorter for the view of 0° than for all other views ($ts[133] \geqslant 2.21$).

2.3. Discussion

The results of Experiment 1A indicated that scene recognition, like judgments of relative direction, could be intrinsic-orientation dependent suggesting that scene recognition relies on spatial memories organized with respect to intrinsic frames of reference (e.g., Mou & McNamara, 2002), rather than viewpoint dependent visual memories (e.g., Diwadkar & McNamara, 1997). To our knowledge, this is the first such demonstration in the spatial memory literature. A possible limitation of this experiment is that different sets of objects were used for each view; it is therefore possible that participants were faster on the test view of 0° because the triplets of objects used in those target scenes were easier to recognize for some reason. Experiment 1B was designed to control for this confound.

3. Experiment 1B

In Experiment 1B, we exchanged the viewing direction and the intrinsic direction in Experiment 1A; that is, participants viewed the layout at the position of 0° and were instructed to learn the locations of objects along the intrinsic axis 315–135°. We used the same test scenes as in Experiment 1A. If scene recognition relies on an intrinsic-orientation dependent spatial representation, then performance should be best at 315°; however, if the results of Experiment 1A were caused by the confounding of materials with test view, performance should still be best at 0°.

3.1. Method

3.1.1. Participants

Twenty university students (10 men, 10 women) participated for the return of monetary compensation.

3.1.2. Materials, design and procedure

The materials, design and procedure were similar to those in Experiment 1A except for the following modifications: In the learning phase, the participant was led to be seated in a chair with 0° as the viewing direction; after the blindfold was removed, the participant was asked to learn the locations of the objects according to the columns in the 315–135° direction, as indicated by the experimenter (e.g., hat–ball; clip–wood; glue–candle; lock).

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

3.2. Results and discussion

Only the responses to the target configurations were analyzed. Response latency of the correct responses was analyzed in an ANOVA with one term for test view (0–315° in 45° steps).

Mean response latency is plotted in Fig. 4 as a function of test view. As shown in the figure, the major finding was that participants were quicker recognizing the target configurations with the view of 315°, which corresponded to the intrinsic direction, than with the view of 0°, which corresponded to the study view. In other words, performance in scene recognition was again better on a novel view (now 315°) than on a familiar view (now 0°).

The main effect of test view was significant, $F(7,133) = 2.79$, $p < .01$, $MSE = 3.29$. Pairwise comparisons showed that response latency was shorter for the view of 315° than for all other views (ts[133] $\geqslant 2.11$) with exception of 45° (t[133] = 1.84) and 135° (t[133] = 1.71), which were significant using one-tailed t-test.

These results eliminate the possibility that the pattern of intrinsic-orientation dependence observed in Experiment 1A was caused by irrelevant visual differences between triplets of objects used at 0° and those used at 315° and suggest that scene recognition relies on spatial memory that is intrinsic-orientation dependent.

As described previously, we hypothesized that the intrinsic-orientation effect observed in Experiments 1A and 1B occurred because all sets of objects used in the intrinsic direction had two objects parallel to the intrinsic reference direction in the mental representation, whereas some sets of objects used in other directions did not (although they did have two objects parallel to the tested viewing direction). We expected that the intrinsic-orientation effect would disappear if the same configuration was tested from different viewpoints because in a specific configuration the interobject spatial relations with respect to the intrinsic reference direction are identical across all viewpoints. Instead a viewpoint dependent pattern would appear because the intrinsic reference direction in the mental representation and the identified intrinsic reference direction in the target scene would need to be aligned when the study view was different from the test view. Experiment 2 was designed to test

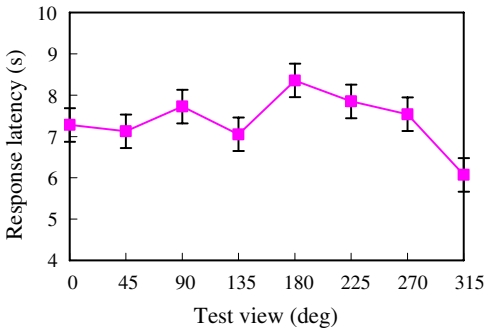

Fig. 4. Response latency in scene recognition as a function of test view in Experiment 1B.

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

this hypothesis. Experiment 2 also allowed us to determine whether the results of Experiments 1A and 1B were caused by participants having learned the layout from an instructed viewpoint. Participants in Experiment 2 were also instructed to learn the layout along a nonegocentric axis, as in Experiments 1A and 1B, but we now expected to find viewpoint dependent performance, because of the nature of the test trials.

4. Experiment 2

In Experiment 2, the study view was 315° and the intrinsic direction was 0°. The test configurations for all test views were identical regardless of the viewpoints. We expected to observe viewpoint dependent performance as in previous studies, such that performance would be best for the study view of 315°.

4.1. Method

4.1.1. Participants

Thirty university students (15 men, 15 women) participated for the return of monetary compensation.

4.1.2. Materials, design and procedure

The materials, design and procedure were similar to those used in Experiment 1A except for the following modifications: six configurations of six objects were produced by removing each object (except the wood; see Fig. 2) from the study configuration one at a time; 48 targets were produced by presenting these six configurations of six objects at eight test viewpoints according to the study configuration; and 48 distractors were produced by presenting the mirror reflections of the targets with respect to the test viewing directions.

4.2. Results and discussion

Only the responses to the targets were analyzed. Response latency for the correct responses was analyzed in an ANOVA with a term for test view (0–315° in 45° steps).

Mean response latency is plotted in Fig. 5 as a function of test view. As shown in the figure, the major findings were these: First, participants were quicker recognizing the target configurations with the view of 315°, which corresponded to the study view, than with the view of 0°, which corresponded to the intrinsic axis. Second, response latency increased with the angular distance (up to 180°) between the test view and the study view.

All of these conclusions were supported by statistical analyses. The overall effect of the test view was significant, $F(7,203) = 6.94$, $p < .001$, $MSE = 1.66$. Pairwise comparisons showed that response latency was shorter for the view of 315° than for all other views ($ts[203] \geqslant 2.18$) with exception of 270° ($t[203] = 1.95$), which was significant using one-tailed t-test. To investigate further the quantitative relation

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

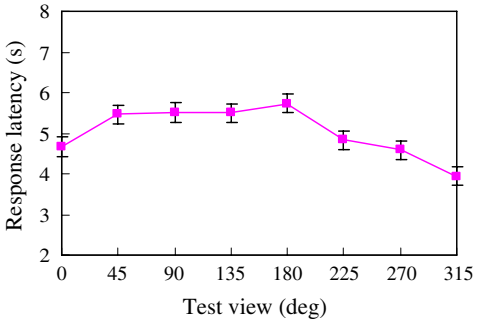

Fig. 5. Response latency in scene recognition as a function of test view in Experiment 2.

between angular distance and performance, we redefined test views in terms of their angular distance from the study view (315° was redefined as 0°; 270° and 0° were redefined as 45°; etc.). Response latency for the correct responses was analyzed in an ANOVA with a term for distance to the study view (0–180° in 45° steps). The linear effect of distance was significant, $t(29) = 5.19$. The quadratic effect of distance was also significant, $t(29) = 2.11$. In particular, there was no significant difference between the distance 135° and the distance 180°, $t(29) = 0.41$. No other polynomial effects were reliable. These results showed that response latency increased with the angular distance between the test view and the study view, up to 135°.

The results of Experiment 2 showed that recognizing the same configuration from different test viewpoints is viewpoint dependent suggesting that participants' viewing direction was represented with respect to the intrinsic frame of reference in spatial memory and also that the intrinsic-orientation dependent pattern observed in Experiments 1A and 1B was caused by the different types of triplets presented at the viewpoints parallel to the intrinsic direction and the viewing direction.

Experiments 1A, 1B, and Experiment 2 supported the hypothesis that the intrinsic-orientation dependent effect was caused by the different triplets and that the viewpoint dependent effect was caused by the alignment between the represented intrinsic reference directions and the identified intrinsic reference direction of the test scene. This conclusion would be strengthened if we could dissociate these two effects in a single experiment. In addition, we needed to deal with the possibility that the different number of objects used in Experiments 1A and 1B (three) and in Experiment 2 (six) caused the different results between these experiments. Experiment 3 was designed to dissociate the effect of the study viewing direction and the effect of the intrinsic reference direction.

5. Experiment 3

In Experiment 3, for one group of participants, the study viewing direction was 315° and the intrinsic direction was 0° and for the other group, the study viewing

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

direction and the intrinsic direction reversed. Intrinsic triplets, among which two objects were presented along the intrinsic direction, and non-intrinsic triplets, among which two objects were presented along the study viewing direction, were presented at all eight test viewpoints. This two-factors design allowed us to investigate the main effects and interaction effect of the intrinsic orientation and the viewing direction in scene recognition.

5.1. Method

5.1.1. Participants

Twenty university students (10 men, 10 women) participated for the return of monetary compensation.

5.1.2. Materials, design and procedure

Six triplets along the axis of 315° were created as follows: hat–ball–candle, hat–ball–lock, glue–candle–clip, glue–candle–hat, clip–wood–lock, clip–wood–candle. Six triplets along the axis of 0° were created as follows: wood–ball–candle, wood–ball–lock, lock–candle–ball, lock–candle–hat, clip–hat–candle, clip–hat–glue. The two sets of triplets were therefore non-overlapping such that triplets along the axis of 0° did not contain any two objects lying along the axis of 315° and the triplets along the axis of 315° did not contain any two objects lying along the axis of 0°. All 12 triplets were tested at all of the 8 test viewpoints (0–315° in step of 45°).

Half of the participants had the study view of 0° and the intrinsic direction of 315° (intrinsic 315 group); the other half had the study view of 315° and the intrinsic direction of 0° (intrinsic 0 group). Participants were randomly assigned to the intrinsic 0 group and the intrinsic 315 group.

One of the primary independent variables was triplets (intrinsic vs. non-intrinsic). The triplets along the intrinsic direction were defined as the intrinsic triplets, whereas the triplets along the study viewing direction were defined as the non-intrinsic triplets. Stimuli were counterbalanced through these two conditions across the two groups of participants. For example, the triplets along the axis of 0° were intrinsic triplets to the intrinsic 0 group but were non-intrinsic triplets to the intrinsic 315 group. The other primary independent variable was the angular distance of the test view relative to the study view. For the intrinsic 0 group, the distance of the test view relative to the study view of 315° was defined in a counterclockwise direction, whereas for the intrinsic 315 group, the distance of the test view relative to the study view of 0° was defined in a clockwise direction so that the test view parallel to the intrinsic direction was defined as 45° from the study view in both groups.

5.2. Results and discussion

Only the responses to the target configurations were analyzed. Response latency of the correct responses was analyzed in mixed ANOVAs with terms for distance of the test views relative to the study view (0–315° in 45° steps), triplets (intrinsic or non-intrinsic) and intrinsic group (intrinsic 315 group or intrinsic 0 group). Distance

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

to the study view and triplets were within-participant and intrinsic group was between-participants.

No effects involving intrinsic group were significant (Fs <1.22). Hence the mean response latency is plotted in Fig. 6 as a function of distance to the study view and triplets. As shown in the figure, the major findings were these: First, participants were quicker recognizing the intrinsic triplets than the non-intrinsic triplets. Second, participants were quicker recognizing the study view (0° in Fig. 6) than other views including the intrinsic axis view (45°) for both types of triplets.

All of these conclusions were supported by statistical analyses. The effect of triplet type was significant, $F(1, 18) = 17.35$, $p < .05$, $MSE = 5.78$. The overall effect of the distance to the study view was significant, $F(7,126) = 4.76$, $p < .001$, $MSE = 5.84$. The interaction between type of triplet and the distance to the study view was not significant, $F(7,126) = 0.46$, $p > .05$, $MSE = 4.08$. Pairwise comparisons showed that response latency was shorter for the distance of 0° than for all other distances (ts[126] $\geqslant 2.06$) except for the distance of 315° ($t[126] = 1.17$). To investigate further the quantitative relation between distance to the study view and performance, we recoded distance from 0° to 180° by averaging the latency at the same distances (e.g., 45° and 315°; 90° and 270°; 135° and 225°). Response latency of the correct responses was analyzed in a mixed ANOVA with terms for distance of the test views relative to the study view (0–180° in 45° steps), triplets (intrinsic or non-intrinsic) and intrinsic group (intrinsic 315 group or intrinsic 0 group). The linear effect of distance was significant, $t(18) = 4.88$. All other polynomial effects were not reliable. These analyses show that response latency increased with the angular distance (up to 180°) between the test view and the study view.

The results of Experiment 3 indicate that the different numbers of objects in the test scenes was not the cause of the different patterns of results in Experiments 1A and 1B and Experiment 2. The results add further support to the hypothesis that the intrinsic-orientation effect was caused by the presence of intrinsic triplets at the test viewpoint parallel to the intrinsic direction but the presence of non-intrinsic

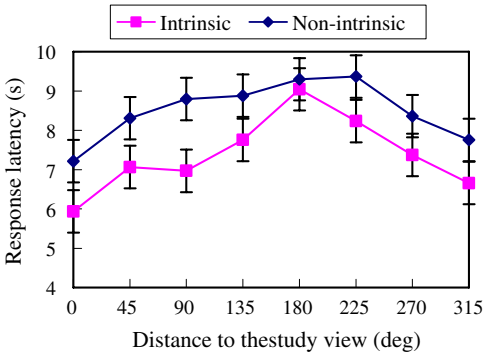

Fig. 6. Response latency in scene recognition as a function of distance between the test view and the study view, and type of test scene in Experiment 3.

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

triplets at the test viewpoint parallel to the study viewing direction. The results also add further support to the hypothesis that the viewpoint dependent effect was caused by the alignment between the intrinsic reference directions in mental representations and in the target scenes.

6. General discussion

The goal of this project was to investigate whether scene recognition relies on visual memory that is viewpoint dependent or spatial memory that is intrinsic-orientation dependent. The findings of the experiments support the latter. The triplets with two objects along the intrinsic direction (intrinsic triplets) were easier to recognize than were the triplets with two objects along the viewing direction (non-intrinsic triplets), even when the intrinsic triplets were presented at a novel viewpoint and the non-intrinsic triplets were presented at a familiar viewpoint in Experiments 1A and 1B. Although the learning view effect was also observed in Experiments 2 and 3, this does not necessarily imply that scene recognition relies on visual memory. Otherwise it is hard to explain why the intrinsic triplets were easier to recognize than non-intrinsic triplets, especially when both were presented at the familiar viewpoint in Experiment 3, and the intrinsic triplets presented at a novel viewpoint were easier to recognize than non-intrinsic triplets presented at the familiar viewpoint in Experiments 1A and 1B. These results, together with the intrinsic-orientation dependent results observed in judgments of relative direction suggest that the inconsistent findings in scene recognition and judgments of relative direction in previous studies were not caused by different tasks. Instead, the different triplets (intrinsic triplets or non-intrinsic triplets) used in the different test views caused the intrinsic-orientation dependent result. When the same configurations of objects were tested at different viewpoints, which has the effect of removing the different intrinsic-orientation effects across different viewpoints, scene recognition appeared to be viewpoint dependent in Experiment 2.

The results from this project elaborate and develop the model of spatial memory and navigation proposed by Mou et al. (2004). According to this model, people represent interobject spatial relations with respect to an intrinsic reference direction and also represent their own viewing direction with respect to the same intrinsic reference direction. When accessing interobject spatial relations, people must identify the intrinsic reference direction in the test scene first. The intrinsic reference direction is easier to identify when the test scene contains at least two objects on an axis parallel to the intrinsic reference direction than when it does not contain such a set of objects. Hence the objects-to-intrinsic-reference-direction relations can explain the intrinsic-orientation dependent result in Experiments 1A and 1B. After people identify the intrinsic reference direction in the test scene, they need to align the represented intrinsic reference direction established at the study viewing direction with the intrinsic reference direction identified in the test scene. No alignment is needed if the intrinsic reference direction identified in the test scene is the same as what people established at their study viewing direction. Hence the self-to-intrinsic-reference-

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

direction relation can explain the viewpoint dependent result in Experiment 2 and in the previous studies (Shelton & McNamara, 2004a, 2004b; Valiquette & McNamara, in press). The results of Experiments 1A and 1B also imply that when identification of the intrinsic reference direction required relatively more effort than alignment of intrinsic reference directions participants were able to recognize the intrinsic triplets presented at a novel viewpoint (0° in 1A and 315° in 1B) more easily than the nonintrinsic triplets presented at the viewing direction (315° in 1A and 0° in 1B).

An alternative explanation of our findings is that people store a visual-spatial snapshot of the study view and an allocentric representation of object-to-object spatial relations (e.g., Valiquette & McNamara, in press). Presentation of a test scene may result in parallel matching processes between the test scene and each of these representations. A test scene containing an intrinsic triplet viewed from the study direction would produce strong signals from both of these matching processes. A test scene containing an intrinsic triplet viewed from a novel direction would be easy to match with the object-to-object representation but requires an alignment process for the visual memory. A test scene containing a nonintrinsic triplet viewed from the study direction would be more difficult to match with the object-to-object representation but could produce a strong signal from the visual memory. Finally, a test scene containing a nonintrinsic triplet viewed from a novel viewpoint would produce weak signals and require normalization processes for both representations to determine whether it is a target. If the accumulation of information from these two matching processes is summative, then independent effects of viewpoint and intrinsic/nonintrinsic could result. We acknowledge that the present findings cannot rule out such an explanation and look forward to seeing new evidence directly proving that the viewpoint dependent scene recognition in this study was really caused by the egocentric visual memory.

To our knowledge, no other contemporary models of spatial memory and navigation can easily explain the present challenging findings. In Sholl's model (e.g., Easton & Sholl, 1995; Holmes & Sholl, 2005; Sholl, 2001; Sholl & Nolin, 1997), the spatial relations among objects are represented in an allocentric object-to-object system. However that allocentric system uses an orientation-independent reference system and a dominant reference direction in this system is established by participants' body front when participants are perceptually engaged with the environment. This model cannot explain how participants can use an intrinsic orientation different from their egocentric front.

Wang and Spelke (2000, 2002) have proposed a model of spatial memory and navigation that consists of both egocentric and allocentric systems. The egocentric system represents and dynamically updates spatial relations between the body and important objects in the surrounding environment. The egocentric system also represents the appearances of familiar landmarks and scenes. These representations are viewpoint-dependent and can be conceived of as visual-spatial "snapshots" of the environment (e.g., Burgess et al., 2004; Diwadkar & McNamara, 1997; Wang & Simons, 1999). The allocentric system represents the geometric shape of the environment (e.g., the shape of a room) but not the spatial relations among objects in the environment. The snapshot representation can well explain the viewpoint dependent

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

scene recognition in this project. However it cannot explain the intrinsic-orientation dependent results. The allocentric system also cannot explain the intrinsic-orientation dependent result because it does not represent the spatial relations among objects.

The present findings may also have important implications for the nature of representations of objects and shapes. People may represent the spatial structure of a shape (object) with respect to the intrinsic reference direction of the shape (object) (e.g., Palmer, 1999; Rock, 1973). People may also represent their own study viewpoint with respect to the same intrinsic reference direction. When they recognize a shape (object), they first need to identify the intrinsic reference direction of the shape (object) and then align the intrinsic reference direction of the test shape (object) with the represented intrinsic reference direction established from the learned viewpoint. If this implication is correct, both intrinsic-orientation dependent and viewpoint dependent recognition are expected. Viewpoint dependent shape (object) recognition has been well documented in the literature (e.g., Cooper, 1975; Jolicoeur, 1988; Tarr & Pinker, 1989, 1990). These findings have been taken as evidence of a viewer-centered representation and the normalization processes used to align viewpoint-specific representations of test stimuli with those represented in memory. As discussed above, in our model, viewpoint dependent performance may not imply that viewer-centered representations are being used, but rather it implies that viewing direction is represented with respect to an object-centered frame. To our knowledge, previous investigations of object and shape recognition have not dissociated viewpoint dependence from intrinsic-orientation dependence. Ongoing projects in our laboratory are testing our model in shape and object recognition paradigms.

In summary, the most important findings from these experiments are the following: First, scene recognition is intrinsic-orientation dependent. This means that an unfamiliar view will be recognized more efficiently if it contains at least two objects aligned with the intrinsic direction selected at the time of learning. We demonstrated this finding by showing that test views that contained two objects aligned with the intrinsic direction were recognized faster than test views that did not contain two objects aligned with the intrinsic direction. Second, scene recognition is also viewpoint dependent. In our model, this result implies that people represent the direction of their original study view with respect to the intrinsic reference direction. All of these findings support and develop the allocentric model of spatial memory and navigation proposed by Mou et al. (2004) with the following properties relevant to scene recognition:

1. Interobject spatial relations are specified with respect to an intrinsic reference direction in the scene.
2. The study viewing direction of the observer is specified with respect to intrinsic reference direction.
3. People identify the intrinsic reference direction in the test scene.
4. People align the intrinsic reference direction in the test scene with the represented intrinsic reference direction specified at the study viewing direction.

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

Acknowledgements

Preparation of this paper and the research reported in it were supported in part by the 973 Program of Chinese Ministry of Science and Technology (2006CB303101) and a grant from the National Natural Science Foundation of China (30470576) to W.M. and National Institute of Mental Health Grant 2-R01-MH57868 to T.P.M. We are grateful to Dr. Gerry T.M. Altmann and two anonymous reviewers for their helpful comments on a previous version of this manuscript. We are also grateful to the Virtual Reality Laboratory of The Institute of Psychology, Chinese Academy of Sciences for using its facility and space in collecting data.

References

Burgess, N., Spiers, H. J., & Paleologou, E. (2004). Orientational manoeuvres in the dark: Dissociating allocentric and egocentric influences on spatial memory. *Cognition, 94*, 149–166.

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

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology, 7*, 20–43.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*, 302–307.

Easton, R. D., & Sholl, M. J. (1995). Object-array structure, frame of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 483–500.

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, & Cognition, 31*, 1069–1087.

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology, 42*, 461–478.

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 representing and processing spatial knowledge. LNAI 1404* (pp. 1–17). Berlin: Springer.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 162–170.

Mou, W., McNamara, T. P., Rump, B., & Xiao, C. (2006). Roles of egocentric and allocentric spatial representations in locomotion and reorientation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1274–1290.

Mou, W., McNamara, T. P., Valiquiette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 142–157.

Owen, C. B., Tang, A., & Xiao, F. (2003). ImageTclAR: A blended script and compiled code development system for augmented reality. STARS2003, The international workshop on software technology for augmented reality systems, Tokyo, Japan.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: The MIT Press.

Rock, I. (1973). *Orientation and form*. New York: Academic Press.

Shelton, A. L., & McNamara, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review, 4*, 102–106.

Shelton, A. L., & McNamara, T. P. (2001). Visual memories from nonvisual experiences. *Psychological Science, 12*, 343–347.

Shelton, A. L., & McNamara, T. P. (2004a). Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 158–170.

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009

Shelton, A. L., & McNamara, T. P. (2004b). Spatial memory and perspective taking. *Memory & Cognition, 32*, 416–426.

Sholl, M. J. (2001). The role of a self-reference system in spatial navigation. In D. Montello (Ed.). Spatial information theory: Foundations of geographic information science (International Conference, COSIT 2001 Proceedings), Lecture Notes in Computer Science (Vol. 2205, pp. 217–232). Berlin: Springer.

Sholl, M. J., & Nolin, T. L. (1997). Orientation specificity in representations of place. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1494–1507.

Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science, 9*, 315–320.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*, 233–282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame?. *Psychological Science 1*, 253–256.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.

Valiquette, C., & McNamara, T. P. (in press). Different mental representations for place recognition and goal localization. *Psychonomic Bulletin & Review*.

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 867–882.

Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition, 70*, 191–210.

Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition, 77*, 215–250.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences, 6*, 376–382.

Please cite this article in press as: Mou, W. et al., Intrinsic frames of reference and ..., *Cognition* (2007), doi:10.1016/j.cognition.2007.04.009