Implicit learning of viewpoint-independent spatial layouts

Taiga Tsuchiai [1,2], Kazumichi Matsumiya [1,2,3], Ichiro Kuriki [1,2,3] and Satoshi Shioiri [1,2,3] *

[1] Graduate School of Information Sciences, Tohoku University, Sendai, Japan
[2] Japan Science and Technology Agency, CREST, Sendai, Japan
[3] Research Institute of Electrical Communication, Tohoku University, Sendai, Japan

We usually perceive things in our surroundings as unchanged despite viewpoint changes caused by self-motion. The visual system therefore must have a function to process objects independently of viewpoint. In this study, we examined whether viewpoint-independent spatial layout can be obtained implicitly. For this purpose, we used a contextual cueing effect, a learning effect of spatial layout in visual search displays known to be an implicit effect. We investigated the transfer of the contextual cueing effect to images from a different viewpoint by using visual search displays of 3D objects. For images from a different viewpoint, the contextual cueing effect was maintained with self-motion but disappeared when the display changed without self-motion. This indicates that there is an implicit learning effect in environment-centered coordinates and suggests that the spatial representation of object layouts can be obtained and updated implicitly. We also showed that binocular disparity plays an important role in the layout representations.

Keywords: viewpoint, implicit learning, contextual cueing effect, 3D, self-motion, environment-centered coordinates, retinal coordinates, spatial updating

INTRODUCTION

Self-motion changes the viewpoint for seeing objects in our surroundings. Despite the occurrence of a viewpoint change, we perceive the objects and the scene around us as unchanged. For stable visual perception, there must be a mechanism for dealing with objects in extra-retinal/physical coordinates (spatial or environment-centered coordinates). Whether the visual system has environmental-centered representations can be examined by investigating whether perception depends on viewpoint. Viewpoint-independent perception would indicate that the visual system has representations that can be dealt in environment-centered coordinates at least functionally.

There have been several studies on the viewpoint dependence/independence of processes for object recognition. Some researchers have shown that object perception is dependent on viewpoint (Ullman, 1989; Bulthoff et al., 1995; Diwadkar and McNamara, 1997; Hayward and Williams, 2000), whereas others have shown that object perception is independent of it (Biederman, 1987; Hummel and Biederman, 1992). Although the question is still open, there could be different processes with differences in viewpoint dependence (Burgund and Marsolek, 2000).

Layout perception is also important not only for moving around the space but also for object perception. We need spatial layout information for getting to a place, searching things, interacting with objects as well as for object perception. Appropriate spatial layouts facilitate object perception in variety of conditions (Loftus and Mackworth, 1978; Shioiri and Ikeda, 1989; Hollingworth and Henderson, 1999; Sanocki, 2003; Bar, 2004; Davenport and Potter, 2004; Sanocki and Sulman, 2009).

The effect of viewpoint in the perception of object layouts is different from that in the perception of objects in terms of potential contribution of self-motion. Viewing location is a critical factor in layout perception because self-motion changes the retinal image of the whole scene in a systematic manner. This is different from object perception, where object motion also causes a change in the retinal image of each object. Simons and Wang reported little effect of viewpoint change due to self-motion on perceiving scenes or object layouts (Simons and Wang, 1998; Wang and Simons, 1999). This contrasts to a clear viewpoint dependency without self-motion (McNamara and Diwadkar, 1997). They suggest that self-motion is a critical factor although self-motion may not be required when rich visual cues are present (Riecke et al., 2007; Mou et al., 2009; Zhang et al., 2011). The visual system likely has representations of scenes and layouts expressed in environment-centered coordinates and compensates for viewpoint changes that are associated with self-motion.

On one hand, several studies suggested that locations of object representations are automatically encoded (Hasher and Zacks, 1984; Ellis, 1990; Pouliot and Gagnon, 2005) while they did not investigate the effect of viewpoint changes. On the other hand, an automatic process has been suggested to update location representations with self-motion (Farrell and Thomson, 1998). Farrell and Thomson reported an interference effect by self-motion on a pointing task without vision. Localizing error for pointing memorized location increased when participants rotated their body but pointed locations as if they had not rotated it, comparing with the condition without self-motion. Since there was no visual information, the difference should be attributed to self-motion. This interference effect indicates that an automatic process updates spatial representations, which the participant could not ignore, although it does not rule out contributions of other processes (Naveh-Benjamin, 1987, 1988; Caldwell and Masson, 2001).

Our question is whether the layout information in environment-centered coordinates can be processed implicitly

and automatically. An implicit process that allows us to memorize scenes and object layouts is expected to be essential in the everyday life, particularly if it works in environment-centered coordinates. To move around and to interact with objects, we should recognize objects fixed in the environment-centered coordinates. There is normally no difficulty in such tasks and we do not need explicit effort: no intention to memorize locations of objects and no consideration of self-motion to update locations of object representations (including ignoring it). Considering the previous studies, we predicted that layout information is collected and updated implicitly and automatically with self-motion.

To examine whether the representation of the layout is implicitly obtained and used from different locations after self-motion (i.e., updated), we adopted the contextual cueing effect. The contextual cueing effect is a learning effect of spatial layout in visual search displays and is known to be an implicit learning effect (Chun and Jiang, 1998). The participants repeatedly searched for a target among distracters in a visual search task in that experiment. A half of the display layouts, that is, the locations of the target and distracters, were unique to each trial, whereas the remaining layouts were repeatedly used throughout the experiment. The contextual cueing effect makes target detection faster in the repeated layouts. This is implicit learning since the participants are unaware of the repetitions and the layouts are not recognized correctly (the selection rate of layouts, as previously reported, is no better than chance).

We conducted experiments to examine whether the contextual cueing effect survives across a viewpoint change caused by participant self-motion. The stimuli were located in a 3D space presented by a virtual reality system, and the effect of the viewpoint change was investigated. After learning layouts from a particular viewpoint, images from a new viewpoint were presented with or without participant self-motion. If the contextual cueing effect is found under conditions where the viewpoint change is caused by the self-motion, we can conclude that an implicit process is responsible for learning the layouts and updating them so that the layout information is represented in environment-centered coordinates, at least functionally. Although little or no contextual cueing effect is reported with viewpoint changes (Chua and Chun, 2003), no study investigate the effect of self-motion.

We also examined the influence of binocular depth cues. Without depth perception, appropriate spatial representations cannot be achieved. No previous study of spatial memory and updating seriously considered the effect of depth perception. More precise 3D information with binocular disparity might aid in building reliable representations with viewpoint changes, although pictorial depth cues of shading, linear perspective, and size changes may be sufficient. Investigation of the effect of depth cues on spatial perception should provide important insights for uses of virtual reality systems.

We use the term "environmental-centered coordinate representation" simply in contrast to "retinal coordinate representation." Environmental-centered coordinate representation here indicates a system that can functionally cope with changes in retinal image due to self-motion, whatever the underlying mechanism is. Two systems are often assumed for spatial layout representations: the allocentric and egocentric systems (Burgess, 2006; Mou et al.,

2006; Wang et al., 2006). We discuss the relationship between our findings and these two systems in Section "General Discussion."

EXPERIMENT 1

In Experiment 1, we investigated the transfer of the contextual cueing effect to a stimulus viewed from a different location.

PARTICIPANTS

The participants were 20 naive males, all of whom had normal or corrected-to-normal visual acuity.

APPARATUS

The participant wore a head-mounted 3D display (Z800 3DVI-SOR, eMargin) that presented stimuli. Participants with glasses had no difficulty in wearing the head-mounted display. The display size was $32° × 24°$ with a resolution of $800 × 600$ pixels. A computer generated stereo images of 3D CG objects on the display. Two chairs were used to set the two viewing positions, based on which the corresponding stimuli were calculated. The experiment was performed in a dark room.
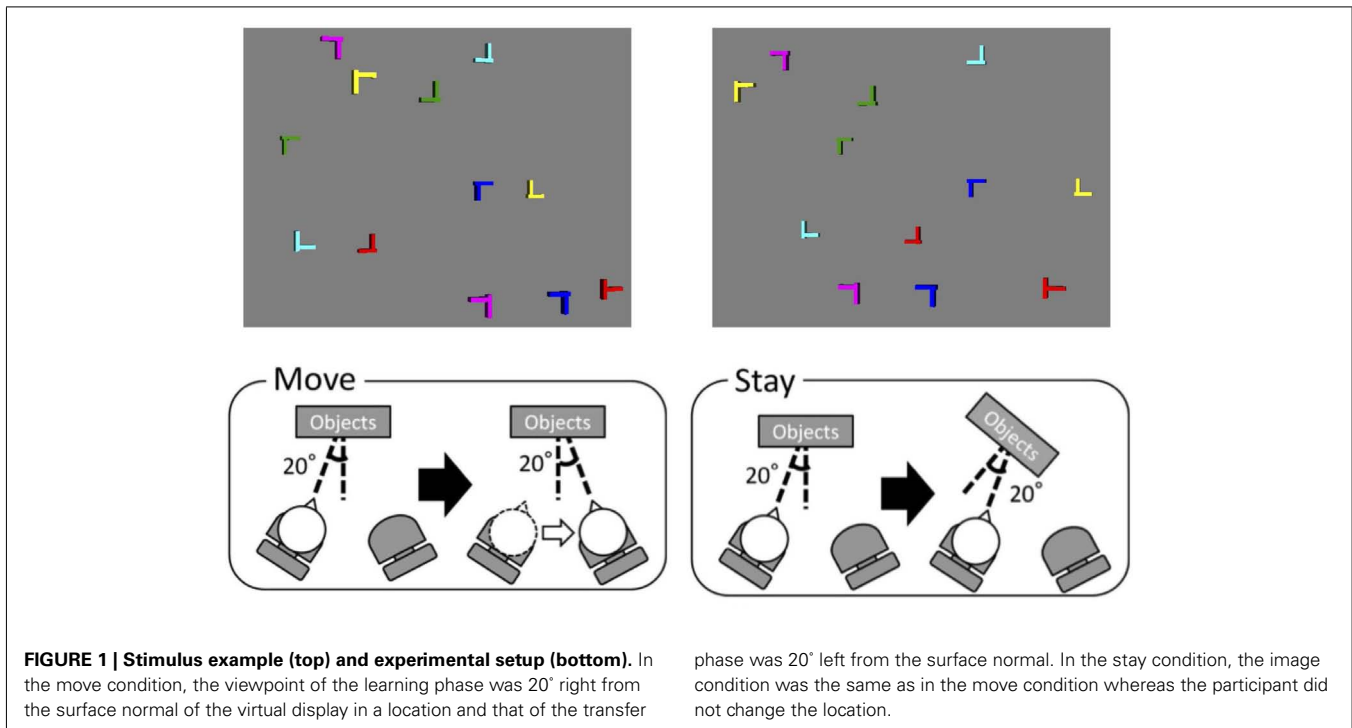
STIMULI

Each stimulus display consisted of 1 T as the target and 11 L's as distractors (**Figure 1**). Each letter object consisted of two square poles and was within a $1.0 cm × 1.0 cm × 0.2 cm$ box. The display size for a view was about $2.0° × 2.0°$. The target orientation was either 90° or 270° (the foot of the T pointed either left or right), and the distractor orientation was either 0°, 90°, 180°, or 270°. The color of the objects was randomly assigned from red, blue, green, yellow, cyan, and magenta. The background was gray. The target and distractors were positioned at randomly chosen intersections of an invisible $6 × 4 × 4$ grid with jitters (about 0.2° on the display) to avoid linear arrangements. The images were drawn on the display assuming a viewing distance of 60 cm to the center of the front surface of the virtual grid.

There were two stimulus conditions: 2D and 3D stimuli. The 3D stimulus was a pair of stereo images calculated for the 3D letters, and the 2D stimulus was the same image presented to each eye. There was no depth information of disparity variations among letters in the 2D stimulus, whereas depth was perceived in the 3D stimulus based on disparity. Pictorial cues of shading, linear perspective, and size changes were in both the 2D and 3D stimuli.

PROCEDURE

A session consisted a learning phase and a transfer phase. In the learning phase, the participant searched for the target in a number of trials, through which display layouts were learned. The participant indicated the target direction (the foot directed either left or right) by pressing one of two buttons. There were two types of layouts as in a typical contextual cueing experiment. One was the repeated layout, which was used once in each block, and the other was a novel layout, which was used only once throughout a session.

At the beginning of each trial, a fixation point was presented for a randomly chosen period between 0.75 and 1.5 s. At the termination of the fixation point, a search display was presented and terminated when the participant responded. The computer measured the reaction time (RT) for target detection. Each block

FIGURE 1 | Stimulus example (top) and experimental setup (bottom). In the move condition, the viewpoint of the learning phase was 20° right from the surface normal of the virtual display in a location and that of the transfer phase was 20° left from the surface normal. In the stay condition, the image condition was the same as in the move condition whereas the participant did not change the location.

consisted of 20 trials: 10 repeated and 10 novel layouts. After 20 learning blocks, two transfer blocks were tested. There were two transfer conditions: the "move" condition and the "stay" condition. In the move condition, the participant moved from one location to the other after training, whereas the participant stayed at the same position in the stay condition. The two chairs were positioned such that the head was directed ±20° from the surface normal of the virtual display (**Figure 1**). All letters faced along the surface normal. The viewpoint in the learning phase was 20° left from the surface normal (see stimulus example in **Figure 1**), and that of the transfer phase in the move condition was 20° right. The participants moved from one chair to the other. There was a 1-min interval between the learning and transfer phases, during which the participants executed the movement. During the movement, the fixation point was shown on the display and no retinal information of self-motion was provided. In the stay condition, the images of the learning and transfer conditions were the same as in the move condition while the participant stood up from and sat down in the same chair (20° left) during the interval between the two phases. The order of the conditions was counterbalanced across participants.

A recognition test was performed at the end of each session to examine whether the participant had memory explicitly retrieved for each repeated layout. In the recognition test, a distractor replaced the target in each repeated layout. Ten repeated layouts from the search experiment were mixed with 10 novel layouts that had not been shown before. For each of 20 layouts, the participant was asked three questions: whether he had seen the layout in the search blocks, where the target was located if the layout was recognized, and his confident rate for the first answer. Since the recognition rate was not higher than chance level in all conditions, we show only the recognition rates below.

RESULTS AND DISCUSSION

Figure 2 shows RT in the learning phase in the move condition. Each data point shows RT as a function of epoch, which is average RT of two blocks consisting of 20 trials. Shorter RT was found for the repeated layouts compared with the novel layouts after epochs of learning replicating the previous results (Chun and Jiang, 1998). **Figure 3** shows the difference in mean RT between novel and repeated layouts, which we define as the contextual cueing effect. The contextual cueing effect in the transfer phase was calculated from the average of the two blocks, and that in the learning phase was calculated from the average of the last two blocks. Results of a one-tailed t-test for the difference between the novel and repeated layout (whether RT for repeated layouts was shorter or not) was shown by asterisks and plus in the RT results: ** for $p < 0.01$, * for $p < 0.05$, and + for $p < 0.1$ (the same in **Figures 5**, **7** and **9**). We used one-tailed test here because contextual cueing effect is well documented to be RT shortening effect while we used two-tailed tests for other cases below.

A difference was found between the learning and transfer phases in the stay condition, whereas little difference was seen in the move condition. A three-way within-participants ANOVA was conducted for each measure: two disparity conditions (with and without disparity) × two phases (learning and transfer) × two self-motion conditions (stay and move). A significant main effect of phase and self-motion was found: $F(1, 19) = 5.51$, $p < 0.05$ for the transfer effect and $F(1, 19) = 18.06$, $P < 0.001$ for the self-motion effect. Significant interaction was found between phase and self-motion: $F(1, 19) = 4.80$, $p < 0.05$. Our primary interest was in transfer effect and the planed comparison was performed in the contextual cueing effect between the learning and transfer phases. A t-test revealed a significant difference in contextual cueing effects between the two phases in the stay condition [$t = 2.11$,

$p < 0.05$ (2D), $t = 2.29$, $p < 0.05$ (3D)], whereas no such difference was found in the move condition [$t = 0.02$, $p = 0.49$ (2D), $t = 0.13$, $p = 0.45$ (3D)]. These results indicate that the effect of viewpoint changes is influenced by self-motion.

Importantly, the result in the stay condition, where the same stimulus as that in the move condition was used without self-motion, showed little, or no contextual cueing effect in the transfer phase, a result consistent with previous studies (Chua and Chun, 2003). The contextual cueing effect in the move condition must rely on the layout representation that can be updated
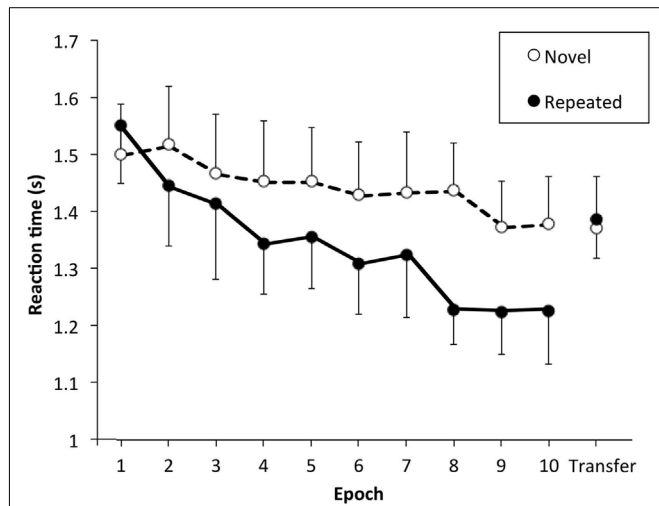

FIGURE 2 | Reaction time as a function of epoch in the stay condition of 3D stimulus shown as a representative. Two blocks are averaged into one epoch. The vertical axis indicates reaction time, and the horizontal axis indicates epoch. Data connected by lines show the results in the learning phase and the rightmost data show the results in the transfer phase. The open circle represents reaction time for the novel configurations, and the solid circle represents that for the repeated configurations. Error bars represent standard error across participants.

by self-motion. This indicates that there is a mechanism to represent objects in environment-centered coordinates or a functionally equivalent mechanism. The result also rules out any interpretations based on stimulus image similarities between the two viewpoints.

For the data collected in the recognition test, a t-test showed that the recognition rate was not significantly different from that expected by chance in all conditions: 49.3% ($t = -0.57$, $p = 0.72$), 51.8% ($t = 1.02$, $p = 0.15$), 49.0% ($t = -0.46$, $p = 0.68$), and 51.3% ($t = 0.72$, $p = 0.23$) for 2D stay, 2D move, 3D stay, and 3D move. In addition, none of the participants reported noticing the repetitions or trying to memorize the configurations. The representation of layouts thus would have been learned implicitly.

EXPERIMENT 2

In Experiment 2, we examined whether the visual system has a retinal representation in addition to an environmental-centered coordinate. If there is only an environmental-centered coordinate system, the contextual cueing effect will vanish in the move condition when the same retinal images are used between the learning and test phases. In other words, no contextual cueing effect was expected when the same retinal image was used at different positions between the learning and transfer phases. However, if the visual system also has a retinal coordinate system for layout learning, the contextual cueing effect will remain after self-motion.

PARTICIPANTS AND PROCEDURE

Sixteen males (including 12 from Experiment 1) participated in Experiment 2. All participants were naive to the purpose of the experiment and had normal or corrected-to-normal visual acuity. The stimuli and procedure were the same as in Experiment 1 with two exceptions. First, the stimulus image in the transfer phase was the same as the image in the learning phase. That is, the image from the original viewpoint, not from the viewpoint at the second location, was used (**Figure 4**). Second, we used only the move condition.
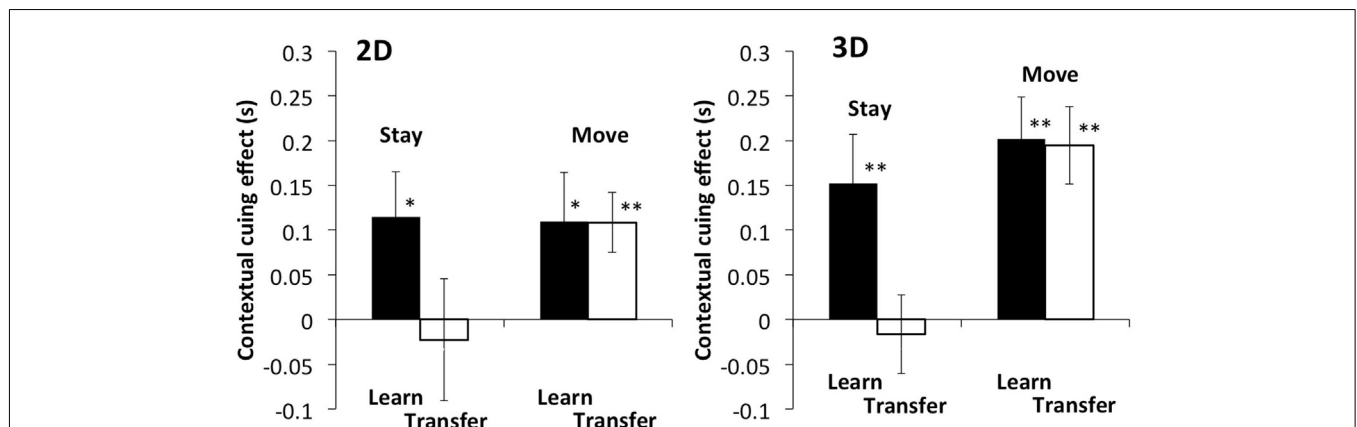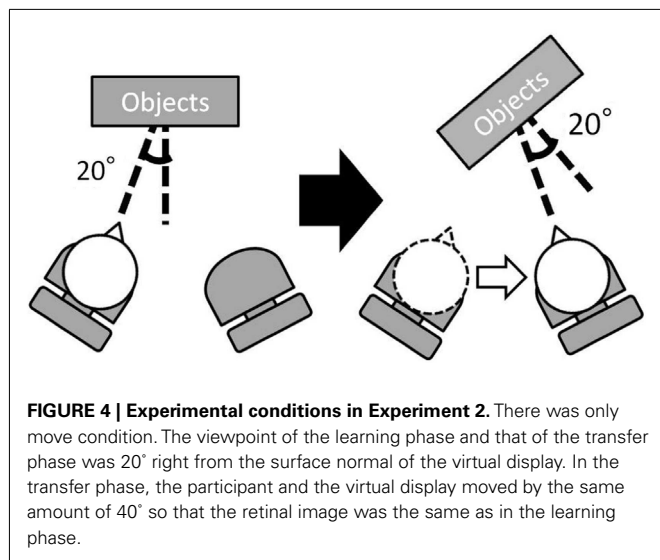

FIGURE 3 | Left and right panels show the results of 2D and 3D stimuli, respectively. The vertical axis indicates the difference in reaction time between the novel and repeated configurations, that is, the contextual cueing effect. The black bar shows the results of the last two epochs of the learning phase, and the white bar shows the results of the

epochs in the transfer phase. Error bars represent standard error across participants. Asterisks and pluses are used to show the results of a one tail t-test for the difference between the novel and repeated layout: ** for $p < 0.01$, * for $p < 0.05$, and + for $p < 0.1$ (the same in **Figures 5**, **7** and **9**).

RESULTS AND DISCUSSION

The results in the learning phase showed contextual cueing effect both in the 2D and 3D conditions as in Experiment 1 (**Figure 5**). In the transfer phase, contextual cueing effect is found also both in the 2D and 3D conditions. A two-way within-participants ANOVA (two disparity conditions × two phases) showed no main effect of disparity [$F(1, 15) = 0.001$, $p = 0.98$] and phases [$F(1, 15) = 0.004$, $p = 0.95$] and no interaction [$F(1, 15) = 0.038$, $p = 0.85$]. A t-test revealed no significant difference in the contextual cueing effect between the learning and transfer phases [$t(15) = 0.17$, $p > 0.1$ for 2D and $t(15) = 0.13$, $p > 0.1$ for 3D]. The contextual cueing effect was observed when the participant moved and the stimulus was unchanged. This suggests that a retinal coordinate system was involved in the contextual cueing effect, in addition to the environment-centered coordinates system.

The recognition test after each session showed that participants recognized the repeated layouts no more than chance [$t(15) = 1.04$, $p > 0.1$ for 2D and $t(15) = 0.59$, $p > 0.1$ for 3D].

FIGURE 4 | Experimental conditions in Experiment 2. There was only move condition. The viewpoint of the learning phase and that of the transfer phase was 20° right from the surface normal of the virtual display. In the transfer phase, the participant and the virtual display moved by the same amount of 40° so that the retinal image was the same as in the learning phase.

Note that 12 participants were from Experiment 1, and this was the second time for them to participate in the contextual cueing experiment. Nonetheless, implicit learning was robustly found in those participants. For them, the contextual cueing effect in the learning phase showed no difference between Experiments 1 and 2 [$t(11) = 0.54$, $p = 0.74$], although RT was slightly shorter in Experiment 2.
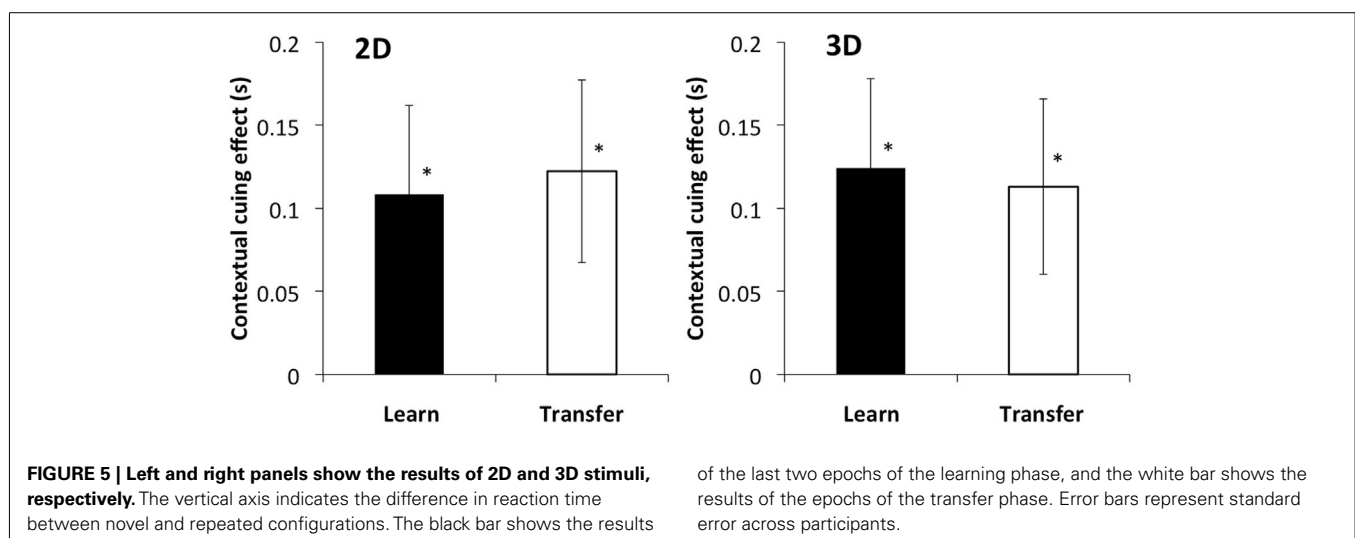
EXPERIMENT 3

In Experiment 1, the contextual cueing effect was found with a viewpoint change of 40°. We examined whether contextual cueing effect transfers to a larger angle, namely, 90°, in Experiment 3. The contextual cueing effect may not transfer to large angle differences because spatial updating has been reported to dependent on viewpoint with large viewpoint changes (Waller and Hodgson, 2006).

PARTICIPANTS AND PROCEDURE

The participants in Experiment 3 were the 12 men who took part in Experiments 1 and 2. All participants were naive to the purpose of the experiment. The stimuli and procedure were the same as in Experiments 1 and 2, except that the viewpoint change was 90° instead of 40° (**Figure 6**). There are two stimulus images in the transfer phase: the retinal image from the new viewpoint as in Experiment 1 (different view images) and the retinal image from the viewpoint of the learning phase (no image change).

RESULTS AND DISCUSSION

The results in the learning phase showed the contextual cueing effect in all conditions (**Figure 7**). A three-way within-participants ANOVA was conducted for the data in the different view image condition as in Experiment 1: two disparity conditions (with and without disparity) × two phases (learning and transfer) × two self-motion conditions (stay and move). A significant main effect of phase was found: $F(1, 11) = 21.11$, $p < 0.001$ while no significant interaction was found between phase and self-motion: $F(1, 11) = 0.82$, $p = 0.39$. This contrasts to the significant interaction found between phase and self-motions in Experiment 1. A t-test for planed comparisons revealed a significant difference in

FIGURE 5 | Left and right panels show the results of 2D and 3D stimuli, respectively. The vertical axis indicates the difference in reaction time between novel and repeated configurations. The black bar shows the results of the last two epochs of the learning phase, and the white bar shows the results of the epochs of the transfer phase. Error bars represent standard error across participants.
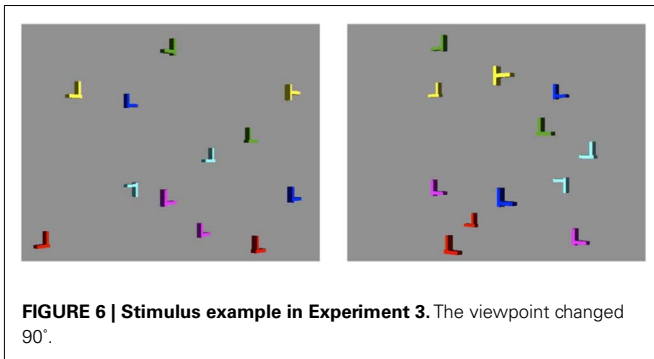
FIGURE 6 | Stimulus example in Experiment 3. The viewpoint changed 90°.

contextual cueing effect between the learning and transfer phases in the 2D stimuli ($t = 2.51$, $p < 0.05$ for the 2D stay condition and $t = 2.77$, $p < 0.05$ for the 2D move condition). The trend of the average data for the 3D stimuli is similar to that for the 2D stimuli although the difference was not significant ($t = 1.80$, $p < 0.1$ for the 3D stay condition and $t = 0.59$, $p = 0.59$ for the 3D move condition). These results suggest that contextual cueing effect disappears in the transfer phase with 2D stimuli and possibly also with 3D stimuli in the condition. There is little transfer of the contextual cueing effect when the viewpoint changed by 90° even with corresponding self-motion.

There is an interesting difference between 2D and 3D stimuli although it is not statistically significant. The contextual cueing effect, the difference between the repeated and novel layouts, for the transfer phase is close to zero for 3D stimuli, whereas the effect for 2D stimuli is below zero. We do not have any interpretation of this pattern of results although this may be related to depth perception and memory of 3D layouts. We leave this issue for future studies.

For the results of the condition with no image change (the same retinal images), a two-way within-participants ANOVA (two disparity conditions × two self-motion conditions) was performed. The result showed no main effect of disparity [$F(1, 11) = 0.26$, $p = 0.62$] and self-motion [$F(1, 11) = 0.79$, $p = 0.39$] and no interaction between disparity and self-motion [$F(1, 11) = 0.00$, $p = 0.99$]. A t-test revealed no significant difference in the contextual cueing effect between the learning and transfer phases ($t(11) = 1.03$, $p = 0.32$ for 2D and $t(11) = 0.64$, $p = 0.54$ for 3D). These results indicate that the impairment by self-motion of 90° is much less for retinal coordinate representations than for environmental-centered representations (Compare top and bottom panels in **Figure 7**).

The recognition rate was not significantly different from chance level ($p > 0.1$ for all conditions): 50.4, 47.1, and 50.4% for the stay, move, and move (same retinal image) conditions with 2D stimuli and 48.3, 49.7, and 52.5% for the stay, move, and move (same retinal image) conditions with 3D stimuli.

EXPERIMENT 4

The previous experiments showed similar results with the 2D and 3D stimuli. This may be surprising if the difference in depth cue is considered because 3D perception is indispensable for layout identification from multiple viewpoints. One possible interpretation for finding no difference is that the depth

information in 2D images was sufficient for the viewpoint-independent contextual cueing effect in previous experiments. Minimizing the effect of monocular depth cues should differentiate the 2D and 3D stimulus conditions because less precise representations in environment-centered coordinates are obtained in the 2D stimuli. In Experiment 4, we used variable sizes of distractors so that size variations in 2D images would confuse the relative depth among stimulus items.

PARTICIPANTS AND PROCEDURE

Twenty new naive participants participated in. All participants had normal or corrected-to-normal visual acuity. The stimuli and procedure were the same as in Experiments 1, except that the size of the distractor letter, L, varied (**Figure 8**). The distractor size varied from −10 to 25% of the size of the original condition. The target size was the same as in the previous experiments.
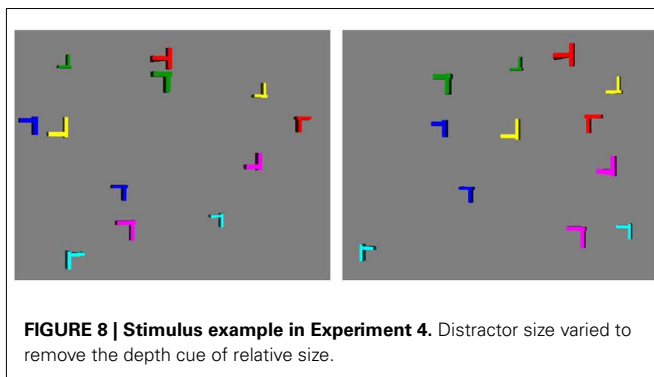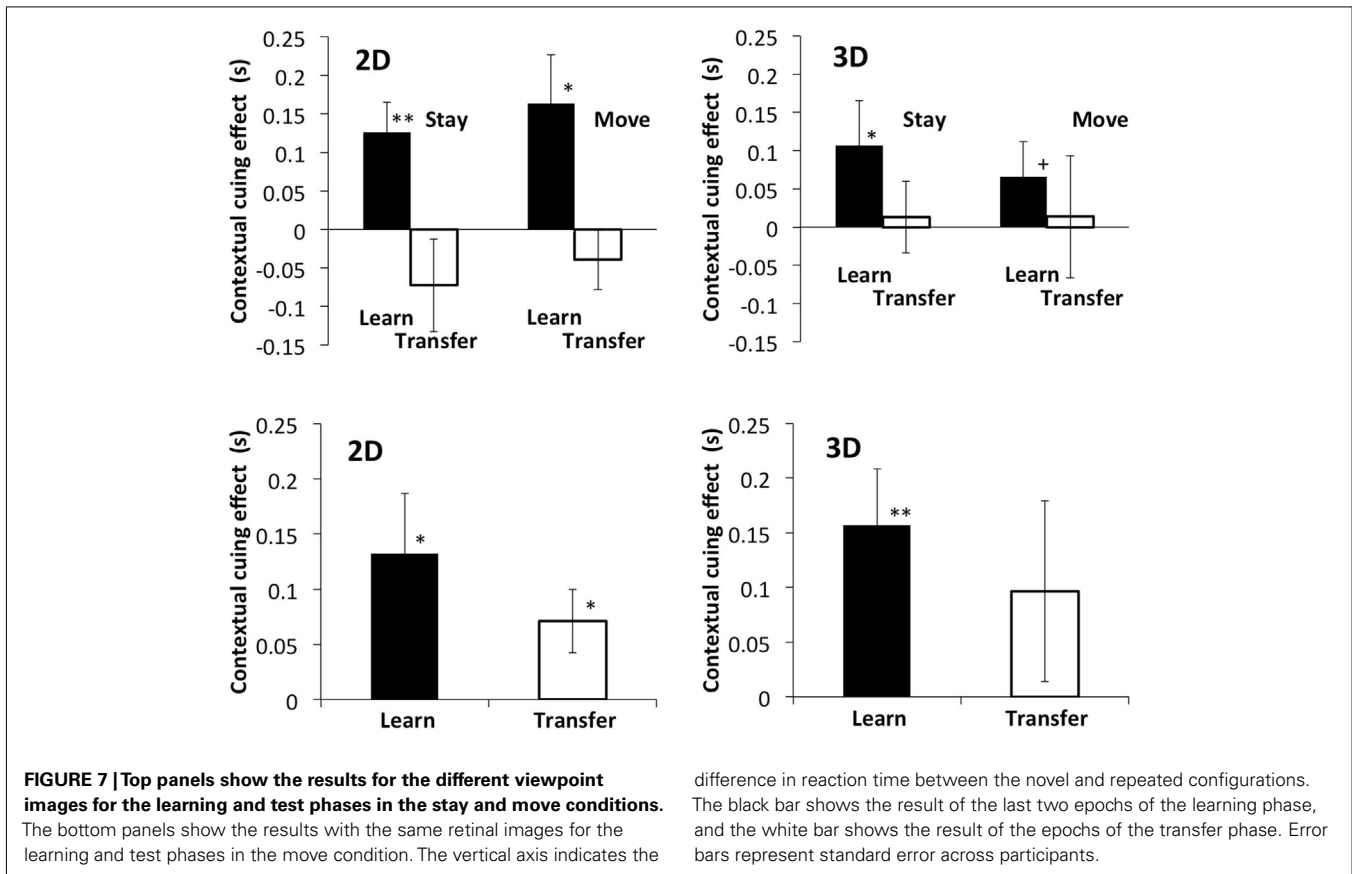
RESULTS AND DISCUSSION

The results revealed a difference in transfer of contextual cueing effect between the 2D and 3D conditions (**Figure 9**). A three-way within-participants ANOVA was conducted: two disparity conditions × two phases × two self-motion conditions. A significant main effects of disparity and phase was found: $F(1, 19) = 7.81$, $p < 0.02$ for disparity and $F(1, 19) = 10.68$, $p < 0.01$ for phase. Significant interaction was found between disparity and phase: $F(1, 19) = 6.48$, $p < 0.02$. A t-test revealed no statistically significant difference in contextual cueing effect between the learning and transfer phases in the 3D move condition [$t(19) = 0.33$, $p = 0.75$] as in Experiment 1. However, the same t-test showed significant difference between the learning and transfer phases in the 2D move condition [$t(19) = 2.22$, $p < 0.05$]. Transfer of the contextual cueing effect disappeared in the 2D stimulus with less monocular depth cues. This indicates that disparity information contributes to build 3D layout representations that can be used with self-motion. The general trend of the results in the stay conditions is similar to that in Experiment 1. However, the t-test revealed no statistically significant difference in contextual cueing effect between the learning and transfer phases in both 2D and 3D conditions [$t(19) = 2.01$, $p = 0.059$ for 2D and $t(19) = 1.22$, $p = 0.235$ for 3D]. The results may imply that the experiment was not sensitive enough to detect the transfer of contextual cueing effect in the condition by some reason.

The recognition rate was not significantly different from chance level ($p > 0.1$ for all conditions): 51.8, 50.5, 51.5, and 52.0% for the 2D stay, 2D move, 3D stay, and 3D move conditions.

GENERAL DISCUSSION

In the transfer phase of Experiment 1, the contextual cueing effect decreased drastically in the stay condition but remained unchanged in the move condition. These results indicate that the viewpoint-independent memory of spatial layouts inevitably needs self-motion as indicated in the previous study by Simons and Wang (Simons and Wang, 1998; Wang and Simons, 1999). The present study revealed that spatial layouts are represented and updated implicitly and automatically in the visual system, showing the transfer of the contextual cueing effect across different viewpoints. The change in layout display did not reduce the contextual

FIGURE 7 | Top panels show the results for the different viewpoint images for the learning and test phases in the stay and move conditions. The bottom panels show the results with the same retinal images for the learning and test phases in the move condition. The vertical axis indicates the difference in reaction time between the novel and repeated configurations. The black bar shows the result of the last two epochs of the learning phase, and the white bar shows the result of the epochs of the transfer phase. Error bars represent standard error across participants.

FIGURE 8 | Stimulus example in Experiment 4. Distractor size varied to remove the depth cue of relative size.
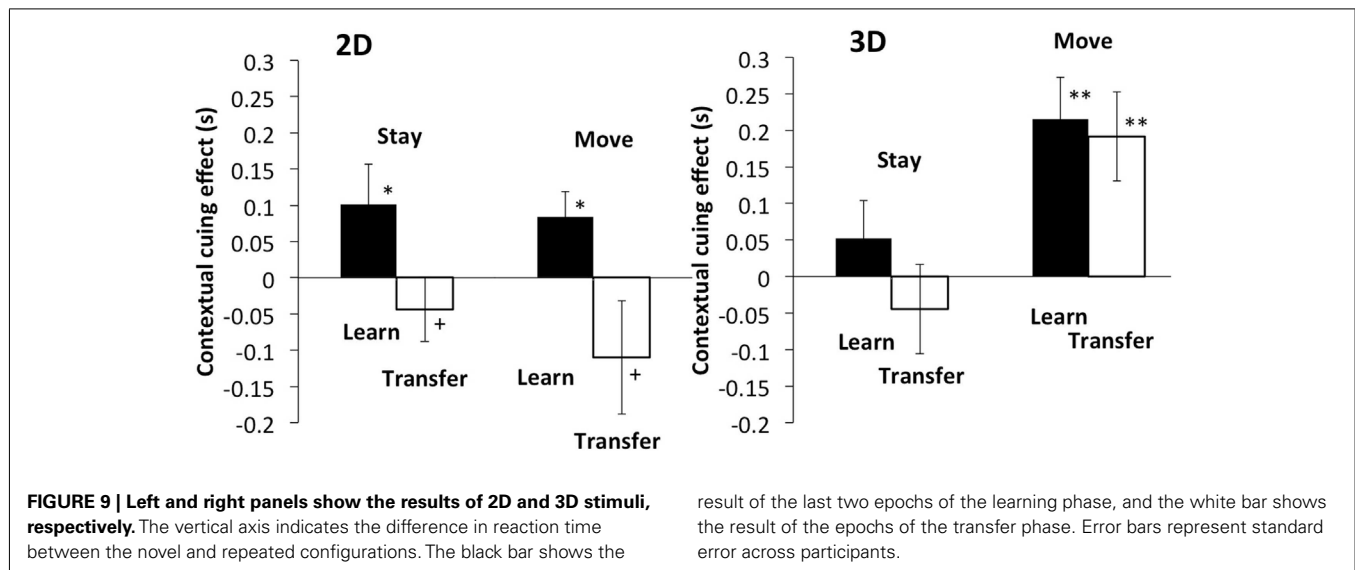
cueing effect when the change was consistent with the viewpoint change due to participant self-motion. The memory should be obtained implicitly since the contextual cueing effect is implicit, as shown in previous studies and our recognition test.

Mou and colleagues pointed out the importance of spatial reference in the spatial memory across self-motion (Mou et al., 2009; Zhang et al., 2011). They conducted similar experiments as that of Simons and Wang and found that cueing a reference direction provided similar or better memory performance without self-motion comparing the one with self-motion. The study revealed that the self-motion related signal is not always the best information for spatial updating and tracking reference direction can be a better cue. This is not necessarily inconsistent with the present results.

Rather, their and our results suggest that different cues are used in different processes for spatial updating.

Experiment 2 suggested that a retinal coordinate system, or snapshots of retinal image, contributes to layout memory. In the condition where the participant moved while the stimulus display did not changed (no image change condition), the contextual cueing effect remained. This is inconsistent with the assumption that there is only an environmental-centered coordinate system. If there were only an environmental-centered coordinate system, the contextual cueing effect should reduce similarly to the effect in the condition where the display changes without self-movement (the stay condition in Experiment 1). However, it should be noted that there is a hint that there is influence of self-motion. That is, Experiment 3 showed some reduction of contextual cueing effect in the no image change condition after self-motion of 90° rotation although the effect is not statistically significant.

An alternative interpretation is possible. Self-motion information may not be used to update layout representations even with self-motion when the participant notices that the layouts are images from the viewpoint of the learning phase. There is no reason to expect reduction of the performance in such occasions. The visual system may have only layout representations in environment-centered coordinates, which may or may not be updated by self-motion dependently on stimulus layouts. This system provides the same function as that of retinal image representation and the present experiments cannot differentiate the two possibilities. Whichever the system is in the visual system,

FIGURE 9 | Left and right panels show the results of 2D and 3D stimuli, respectively. The vertical axis indicates the difference in reaction time between the novel and repeated configurations. The black bar shows the result of the last two epochs of the learning phase, and the white bar shows the result of the epochs of the transfer phase. Error bars represent standard error across participants.

Experiment 2 led us to conclude that the visual system can access to retinal coordinate images.

In Experiment 3, we found that contextual cueing effect cannot survive with self-motion of as large as 90°, but that the retinal coordinate system was influenced much less. This indicates that the implicit memory system assessed in the present study has a limitation for self-motion as previously pointed out for automatic updating (Waller and Hodgson, 2006). Note that richer visual information and/or visual motion during self-motion potentially expands the limitation of viewpoint-independent processing (Riecke et al., 2007; Mou et al., 2009; Zhang et al., 2011).

In Experiment 4, we found that the disparity cue play a critical role in viewpoint-independent contextual cueing effect under the condition, where pictorial depth cues do not provide sufficient depth information. To obtain representations in environment-centered coordinates, depth perception is necessary. The present results showed that manipulation of depth cues influenced the viewpoint-independent effect. Clearly, the implicit viewpoint-independent effect is related to depth perception. This is not surprising but shows importance of 3D perception to investigate the memory of spatial layouts.

For spatial layout representations, two systems are often assumed: the allocentric and egocentric systems (Burgess, 2006; Mou et al., 2006; Wang et al., 2006). Apparently researchers agree that both systems contribute to spatial memory and updating with self-motion (Waller, 2006; Waller and Hodgson, 2006; Sargent et al., 2008). In general, studies on the field of spatial memory focus on much larger scale of spatial layouts than those used in the present study. However, there is no reason to believe that the present results and experimental results with larger scale layouts are processed in different mechanisms. We discuss possible relationship between our results and the allocentric and egocentric systems. On one hand, there are two facts that suggest that the implicit learning/memory phenomenon we found is not related to the allocentric system. First, the allocentric system is suggested to have representations that can be used from any viewpoint (e.g., Waller and Hodgson, 2006). As shown in Experiment 3, the

contextual cueing effect cannot be maintained or updated with a rotation of 90°. Second, a salient feature that can be used as a reference appears to be critical for memorization and retrieval in the allocentric system (Mou et al., 2009; Zhang et al., 2011). No salient feature contributed to our experiments. The layouts consisted of randomly arranged items and the recognition test revealed that there were no memorable salient features.

On the other hand, we think that the implicit learning effect we found is perhaps related more closely to the egocentric system by following reasons. First, the egocentric system is usually characterized by its requirement of self-motion (Wang and Spelke, 2002; Burgess et al., 2004; Rump and McNamara, 2007). The spatial memory system pertaining to the present experiments also required self-motion to update memorized layouts. Second, the egocentric memory is reported to be viewpoint variant for large viewpoint changes (Waller and Hodgson, 2006). We found contextual cueing effect with the 40° viewpoint change, but not with 90°. While the angle of 90° is smaller than that of 135° reported by Waller and Hodgson, the difference is likely attributed to the difference in stimulus and task between the two studies. The experiment of Waller and Hodgson used objects in a room and pointing objects based on intentionally memorized map of the room without visual information. It is not likely that the participant obtain the whole 3D structure of the layout in our experiments because image from only one view was presented. We, therefore, did not expect quantitative similarity in viewpoint dependency between their and our experiments. Based on these discussions, we suggest that the contextual cueing effect we found reflects a characteristic of the egocentric spatial memory system.

There are, however, two issues related the presumption that the egocentric memory system is responsible for contextual cueing effect we found. First, some of the previous studies have suggested that the egocentric memory system is not automatic and requires attention and efforts (Wang and Brockmole, 2003a,b; Wang et al., 2006). This appears to contradict to the implicit effect of contextual cueing. The difference in experimental technique could explain the apparent contradiction. We measured effect on

layouts to which the participants did not have to devote explicit attention. This technique allowed us to investigate the effect of stimuli that are not given attention. In contrast, previous studies of spatial memory measured the effect on stimuli to which the participants had to devote explicit attention. In such experimental conditions, attention very likely influences measurements, in addition to the possible implicit process. Consequently, both explicit and implicit processes could influence the results. The previous studies perhaps obtained results that indicate the influence of attention-requiring system. However, such results do not exclude the possibility that an implicit process also contributes to the performance.

Second, updating of the egocentric process is assumed to depend on the number of objects to memorize by the following reason (Hodgson and Waller, 2006; Wang et al., 2006). A vector expresses the location of an object in an egocentric coordinates. The number of the vectors to update increases with the number of objects in order to maintain the location information during body movements. Therefore, more object processing is required to update locations of more number of objects. This assumption contradicts to the fact that little limitation of information capacity for the contextual cueing effect. The number of layouts learned is often more than 10 and the number of distractors in each layout is usually also more than 10. Spatial updating of layouts learned implicitly is expected to be independent of object numbers. If the body location is updated in environment-centered coordinates, as assumed for allocentric spatial representations, no effect of object number is expected. However we do not discuss this issue in details because we see no clear support for the effect of object number

in spatial memory: Wang et al. (2006) showed effect of object numbers while Hodgson and Waller (2006) did not.

Based on the above discussion we claim that the visual system has an ability to obtain environmental-centered representations and to update them implicitly and automatically with self-motion. Such representations are important to move around and to interact with objects without specific efforts in everyday life. The finding of a retinal representation system in our study is also an important implication for another aspect of spatial perception. A viewpoint specific system has been suggested to play a non-trivial role in spatial perception (Hermer and Spelke, 1996; Lee and Spelke, 2010). When a participant looses egocentric location in the environment (i.e., in environment-centered coordinates), he or she uses visual information viewed from a particular point (snapshots) to estimate his or her own location (Wang and Spelke, 2002). For the estimation, the retinal images learned implicitly can be used. This suggests that the underlying mechanism of the contextual cueing effect in a specific viewpoint is as important as that of a viewpoint-independent, egocentric system for spatial perception while moving in one's surroundings.

ACKNOWLEDGMENTS

REFERENCES

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115–147.

Bulthoff, H. H., Edelman, S. Y., and Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cereb. Cortex* 5, 247–260.

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends Cogn. Sci. (Regul. Ed.)* 10, 551–557.

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

Burgund, E. D., and Marsolek, C. J. (2000). Viewpoint-invariant and viewpoint-dependent object recognition in dissociable neural subsystems. *Psychon. Bull. Rev.* 7, 480–489.

Caldwell, J. I., and Masson, M. E. (2001). Conscious and unconscious influences of memory for object location. *Mem. Cognit.* 29, 285–295.

Chua, K. P., and Chun, M. M. (2003). Implicit scene learning is viewpoint

dependent. *Percept. Psychophys.* 65, 72–80.

Chun, M. M., and Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* 36, 28–71.

Davenport, J. L., and Potter, M. C. (2004). Scene consistency in object and background perception. *Psychol. Sci.* 15, 559–564.

Diwadkar, V. A., and McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychol. Sci.* 8, 302–307.

Ellis, N. R. (1990). Is memory for spatial location automatically encoded? *Mem. Cognit.* 18, 584–592.

Farrell, M. J., and Thomson, J. A. (1998). Automatic spatial updating during locomotion without vision. *Q. J. Exp. Psychol. A.* 51, 637–654.

Hasher, L., and Zacks, R. T. (1984). Automatic processing of fundamental information: the case of frequency of occurrence. *Am. Psychol.* 39, 1372–1388.

Hayward, W. G., and Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychol. Sci.* 11, 7–12.

Hermer, L., and Spelke, E. (1996). Modularity and development: the case of spatial reorientation. *Cognition* 61, 195–232.

Hodgson, E., and Waller, D. (2006). Lack of set size effects in spatial updating: evidence for offline updating. *J. Exp. Psychol. Learn Mem. Cogn.* 32, 854–866.

Hollingworth, A., and Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychol. (Amst.)* 102, 319–343.

Hummel, J. E., and Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* 99, 480–517.

Lee, S. A., and Spelke, E. S. (2010). Two systems of spatial representation underlying navigation. *Exp. Brain Res.* 206, 179–188.

Loftus, G. R., and Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 565–572.

McNamara, T. P., and Diwadkar, V. A. (1997). Symmetry and asymmetry of human spatial memory. *Cogn. Psychol.* 34, 160–190.

Mou, W., McNamara, T. P., Rump, B., and Xiao, C. L. (2006). Roles of egocentric and allocentric spatial representations in locomotion and reorientation. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1274–1290.

Mou, W., Zhang, H., and McNamara, T. P. (2009). Novel-view scene recognition relies on identifying spatial reference directions. *Cognition* 111, 175–186.

Naveh-Benjamin, M. (1987). Coding of spatial location information: an automatic process? *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 595–605.

Naveh-Benjamin, M. (1988). Recognition memory of spatial location information: another failure to support automaticity. *Mem. Cognit.* 16, 437–445.

Pouliot, S., and Gagnon, S. (2005). Is egocentric space automatically encoded? *Acta Psychol. (Amst.)* 118, 193–210.

Riecke, B. E., Cunningham, D. W., and Bulthoff, H. H. (2007). Spatial updating in virtual reality: the sufficiency of visual information. *Psychol. Res.* 71, 298–313.

Rump, B., and McNamara, T. P. (2007). "Updating in models of spatial memory," in *Spatial Cognition V:*

Reasoning, Action, Interaction, eds T. Barkowsky, M. Knauff, G. Ligozat, and D. R. Montello (Berlin: Springer-Verlag), 249–269.

Sanocki, T. (2003). Representation and perception of scenic layout. *Cogn. Psychol.* 47, 43–86.

Sanocki, T., and Sulman, N. (2009). Priming of simple and complex scene layout: rapid function from the intermediate level. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 735–749.

Sargent, J., Dopkins, S., Philbeck, J., and Modarres, R. (2008). Spatial memory during progressive disorientation. *J. Exp. Psychol. Learn Mem. Cogn.* 34, 602–615.

Shioiri, S., and Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception* 18, 347–361.

Simons, D. J., and Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychol. Sci.* 9, 315–320.

Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition* 32, 193–254.

Waller, D. (2006). Egocentric and nonegocentric coding in memory for spatial layout: evidence from scene recognition. *Mem. Cognit.* 34, 491–504.

Waller, D., and Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *J. Exp. Psychol. Learn Mem. Cogn.* 32, 867–882.

Wang, R. F., and Brockmole, J. R. (2003a). Human navigation in nested environments. *J. Exp. Psychol. Learn Mem. Cogn.* 29, 398–404.

Wang, R. F., and Brockmole, J. R. (2003b). Simultaneous spatial updating in nested environments. *Psychon. Bull. Rev.* 10, 981–986.

Wang, R. F., Crowell, J. A., Simons, D. J., Irwin, D. E., Kramer, A. F., Ambinder, M. S., Thomas, L. E., Gosney, J. L., Levinthal, B. R., and Hsieh, B. B. (2006). Spatial updating relies on an egocentric representation of space: effects of the number of objects. *Psychon. Bull. Rev.* 13, 281–286.

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

Wang, R. F., and Spelke, E. (2002). Human spatial representation: insights from animals. *Trends Cogn. Sci. (Regul. Ed.)* 6, 376.

Zhang, H., Mou, W., and McNamara, T. P. (2011). Spatial updating according to a fixed reference direction of a briefly viewed layout. *Cognition* 119, 419–429.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.