Dynamic visual information facilitates object recognition from novel viewpoints

Wataru Teramoto
Max Planck Institute for Biological Cybernetics, Germany, &
Research Institute of Electrical Communication,
Tohoku University, Japan

Bernhard E. Riecke
Max Planck Institute for Biological Cybernetics, Germany, &
Simon Fraser University, Canada

Normally, people have difficulties recognizing objects from novel as compared to learned views, resulting in increased reaction times and errors. Recent studies showed, however, that this "view-dependency" can be reduced or even completely eliminated when novel views result from observer's movements instead of object movements. This observer movement benefit was previously attributed to extra-retinal (physical motion) cues. In two experiments, we demonstrate that dynamic visual information (that would normally accompany observer's movements) can provide a similar benefit and thus a potential alternative explanation. Participants performed sequential matching tasks for Shepard–Metzler-like objects presented via head-mounted display. As predicted by the literature, object recognition performance improved when view changes (45° or 90°) resulted from active observer movements around the object instead of object movements. Unexpectedly, however, merely providing dynamic visual information depicting the viewpoint change showed an equal benefit, despite the lack of any extra-retinal/physical self-motion cues. Moreover, visually simulated rotations of the table and hidden target object (table movement condition) yielded similar performance benefits as simulated viewpoint changes (scene movement condition). These findings challenge the prevailing notion that extra-retinal (physical motion) cues are required for facilitating object recognition from novel viewpoints, and highlight the importance of dynamic visual cues, which have previously received little attention.

Keywords: object recognition, memory, motion-3D, visual cognition

Citation: Teramoto, W., & Riecke, B. E. (2010). Dynamic visual information facilitates object recognition from novel viewpoints. *Journal of Vision, 10*(13):11, 1–13, http://www.journalofvision.org/content/10/13/11, doi:10.1167/10.13.11.

Introduction

Even though the appearance of 3D objects can change rather drastically with changing viewpoints, we have surprisingly little difficulties in identifying and recognizing objects from novel viewpoints in our everyday life. How people manage to perceive and recognize 3D objects (almost) irrespective of the viewpoint has been a fundamental question for researchers of a wide range of research fields such as psychology, neuroscience, and computer science. Several theories about the processes and representations underlying object recognition have been proposed (e.g., Bülthoff, Edelman, & Tarr, 1995; Peissig & Tarr, 2007). While some theories assume 3D view-invariant representations such as geons ("geon structural description", Biederman, 1987; Hummel & Biederman, 1992) others assume viewpoint-specific 3D representations in conjunction with normalization (Ullman, 1989) or multiple viewpoint-dependent 2D object representations with some transformation/interpolation processes in the brain (Bülthoff et al., 1995; Hayward & Williams, 2000; Tarr & Pinker, 1989; Tarr, Williams, Hayward, & Gauthier, 1998). The predominant approach to assess the validity of these theories has been to measure the degree to which view changes affect object recognition performance, with the results ranging from strong viewpoint dependency (increased recognition times and/or errors for novel views) to viewpoint independency, where object recognition performance is independent of the view change (Bülthoff et al., 1995; Peissig & Tarr, 2007).

Simons, Wang, and Roddenberry (2002), however, criticized these prevailing theories because they did not take into account that there are, in fact, two different possibilities for achieving changing views of an object. One is to rotate an object in front of a stationary observer (henceforth called "object-movement" condition), and the other is to move an observer her/himself around a stationary object ("observer-movement" condition). Critically, almost all of the previous studies used only the former (object-movement) procedure. Simons et al. have argued that this might be an "unfair" comparison, and that previous work might have been biased accordingly. This motivated them to examine the difference between object- and observer-movement explicitly in a real-world study. That is, instead of presenting objects on a computer screen or on paper as is customary in the literature, they used a physical environment and physical objects so that one

doi: 10.1167/10.13.11 Received January 5, 2009; published November 12, 2010 ISSN 1534-7362 © ARVO

could physically move around the objects to yield novel views of the objects. They used a sequential matching technique in which two objects were presented successively with a few seconds inter-stimulus interval and the observers judged whether the two objects were identical or not when the second object was presented. In the interval between presentation of the first and the second object, the view of the first object was changed by either object movement or observer movement: In the **object movement condition**, observers remained seated in the same position and thus maintained the same viewpoint while the orientation of the object was changed by 40°. In the **observer movement condition**, observers physically moved to a new viewpoint that was offset by 40° from the initial viewpoint, while the orientation of the object remained unchanged. The results showed that view-dependency was completely eliminated when novel views resulted from observer-movements. The object movement condition, however, resulted in the well-known view-dependency as expected—that is, participants' object recognition performance dropped when the novel view resulted from object motion. Simons and colleagues suggest that the observer's movement helps to recognize objects in the sense that the representation of the object was automatically updated to correspond to the observer's current perspective based on self-motion information. Recently, Zhao, Zhou, Mou, and Hayward (2007) also showed an advantage for the observer movement condition on object recognition performance in a virtual environment using a head-mounted display (HMD). Note, however, that the observer's movement allowed only for "partial" elimination of the effect of viewpoint change on object recognition at 50° of angular disparity. Moreover, 90° disparity failed to show any benefit for observer movement over object movement.

Here, we pursued two main goals: First to investigate whether the advantage of observer movements does indeed vanish for larger (90°) disparities as suggested by Zhao et al. (2007). Second, and more importantly, to investigate where exactly the advantage of observer movements originates from. That is, the current studies were designed to assess which kind of information related to observer's movement is crucial for this phenomenon: In principle, observer's movement can provide richer information for object recognition than object movement in two fundamentally different ways.

First, how much a perspective is transformed is explicitly provided in the observer movement condition but not in the object movement condition. That is, when observers physically move, they can use visual, vestibular and proprioceptive signals to compute the magnitude of the perspective change, even if some specialized mechanisms such as automatic spatial updating are not assumed. For example, Christou, Tjan, and Bülthoff (2003) showed that both implicit (visual background of a room simulated in virtual reality) and explicit (an arrow indicating the initial perspective) indications of perspective change of an object resulted in better performance on subsequent shape recognition. Second, observers in the observer movement condition actively cause perspective changes while they do not in an object movement condition. It is well known that active control can improve various performances. For example, learning that involved active manual exploration of objects (Harman, Humphrey, & Goodale, 1999; James et al., 2002) and virtual environments (Christou & Bülthoff, 1999) improved subsequent recognition performance.

For this reason, Simons et al. (2002) investigated how some of these factors contributed to the advantage of the observer movement condition in their two follow-up experiments. In their main experiments described above, there was an experimenter and a computer monitor visible right behind the table (or the object), such that observers received different views of the room depending on the viewpoint they took. That is, observers saw the same, unchanged visual background during the task in the object movement condition, while they saw a different visual background when they moved to the new viewpoint. Therefore, they designed one of their follow-up experiments to test whether this difference in the visual information available for viewpoint changes had an effect on the recognition performance. Simons et al. used digital photographs of each object taken from the actual viewing positions, instead of actually rotating the table or the observer. If the different views of the room background provided useful information about the magnitude of the viewpoint change, only the presentation of the photos with different visual backgrounds (even if the viewing position was actually not changed) should have facilitated the object recognition performance. The results revealed that presentation of the visual background alone did not improve recognition performance. This result is inconsistent with the findings of Christou et al. (2003) described above. Simons et al. note that this inconsistency might be attributed to the difference in saliency of the visual background—the visual background in Simons et al. was less salient. In the other follow-up experiment, they used the same procedure as in their main experiments, but with a uniformly colored visual background. If the visual background was critical, they should not have found any difference between the observer and the object movement conditions. The results showed that the observer movement condition was still better than the object movement condition. Thus, Simons et al. showed that extra-retinal information (as compared to visual information) was more crucial for facilitating object recognition. However, Simons et al. (2002, and Christou et al., 2003, as well) only tested the effect of "static" visual information on object recognition. Visual signals available for viewpoint change can in fact be subdivided into two types: static visual information (e.g., visual landmarks or room geometry uniquely defining one's viewpoint) and dynamic visual information (e.g., optical flow cues originating from the observer and/or object movement, such as in our case the visual movement of the whole visual scene and/or the

table). Note that previous studies only investigated the influence of static, but not dynamic visual information on object recognition from novel viewpoints.

The goal of the current study was to close this gap and investigate whether dynamic visual information might provide a benefit for the object recognition from novel viewpoints, potentially similar to the benefit observed from spatial updating from dynamic non-visual (bodily) movement. It is important to note that dynamic visual information includes the amount of perspective change, regardless of whether the change is caused by object motion or observer motion. That is, it is conceivable that the mere presentation of dynamic visual information about the orientation change of an object, but not the viewpoint change of an observer, might be sufficient to facilitate subsequent object recognition. To test this hypothesis, we compared three conditions with each other using a sequential matching task similar to the one used by Simons et al. (2002). The first was an observer movement condition, which was the same as the one tested in the previous studies. In this condition, the observer moved to a next viewpoint during the time interval between the presentation of the first and the second object. The second was a scene movement condition where the whole visual scene was rotated without any actual observer movement or self-motion perception. In other words, only visually simulated viewpoint change was presented in this condition. The third was a table movement condition where only the table and things on the table was rotated while the visual background remained stationary. Only in the first (observer movement) condition did observers actively cause the viewpoint change. Thus, if extra-retinal or active control information about the perspective change was essential, the observer movement condition should outperform the other two conditions.

If dynamic visual information about viewpoint change of the observer contributes to object recognition from novel viewpoints, performance in the scene movement condition should be better than for table movement condition and potentially even approach observer movement performance if physical motion would not be as essential as previously thought. If the dynamic presentation of the amount of perspective change was most important for this phenomenon, regardless of whether the change is caused by orientation change of an object or viewpoint change of an observer, the three conditions should yield similar results.

General methods

Participants

Twelve people between 21 and 32 years of age (6 females and 6 males) participated in both Experiments 1 and 2, and were paid for each hour of participation. All participants had normal or corrected-to-normal vision and were naïve to the purposes of the experiments. Informed consent was obtained from each participant before the experiments.

Apparatus and materials

The experiments were conducted using an eMagin Z800 3D Visor HMD (eMagin Corporation) that was tracked by 16 Vicon MX 13 motion capture cameras (Vicon Motion Systems Ltd., California, temporal resolution up to 484 Hz). The HMD supplied two identical images to both eyes at a resolution of 800 × 600 pixels and a field of view of 40° diagonally for each eye (32° × 24°, refresh rate of 60 Hz in mono mode). The virtual space was rendered using veLua, a custom designed open source Virtual Reality communications and rendering library (http://velib.kyb.mpg.de/veLua/index.html). The 3D model of the virtual environment was developed using 3ds Max (Autodesk Inc., California). The virtual environment depicted in the HMD changed according to the participants' head movements, providing motion parallax and structure-from-motion cues.

As shown in Figures 1 and 2, the simulated room was a cylindrical room whose wall was textured with a gray marble pattern. In the center of the simulated room, an upright cylinder with a radius of 10.5 cm was positioned on a round table with a radius of 30 cm and a height of 110 cm. The height of the cylinder was changeable so that participant's viewing height was aligned with the center of the viewing window. The outside of the cylinder was textured with a 50% black–white random-dot pattern and the inside was colored black. The cylinder had eight

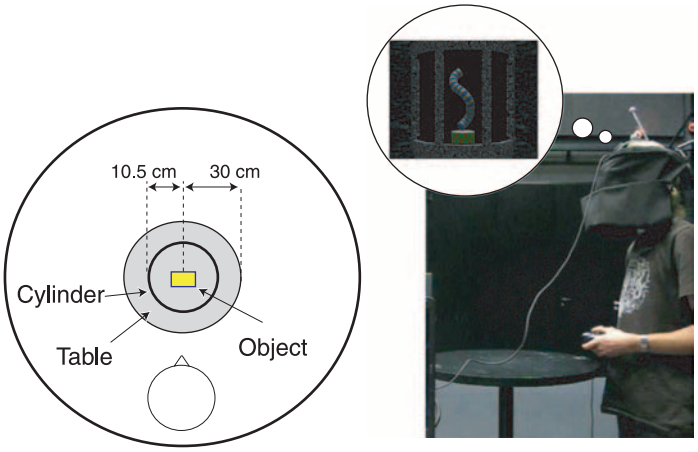

Figure 1. Left: Top-down sketch of the experimental setup. Right: Picture of a participant with input device (gamepad) and wearing the head-mounted display (HMD) and tracking helmet. Note that the HMD displays a scene that is aligned with the physical table in front of which participants are positioned.

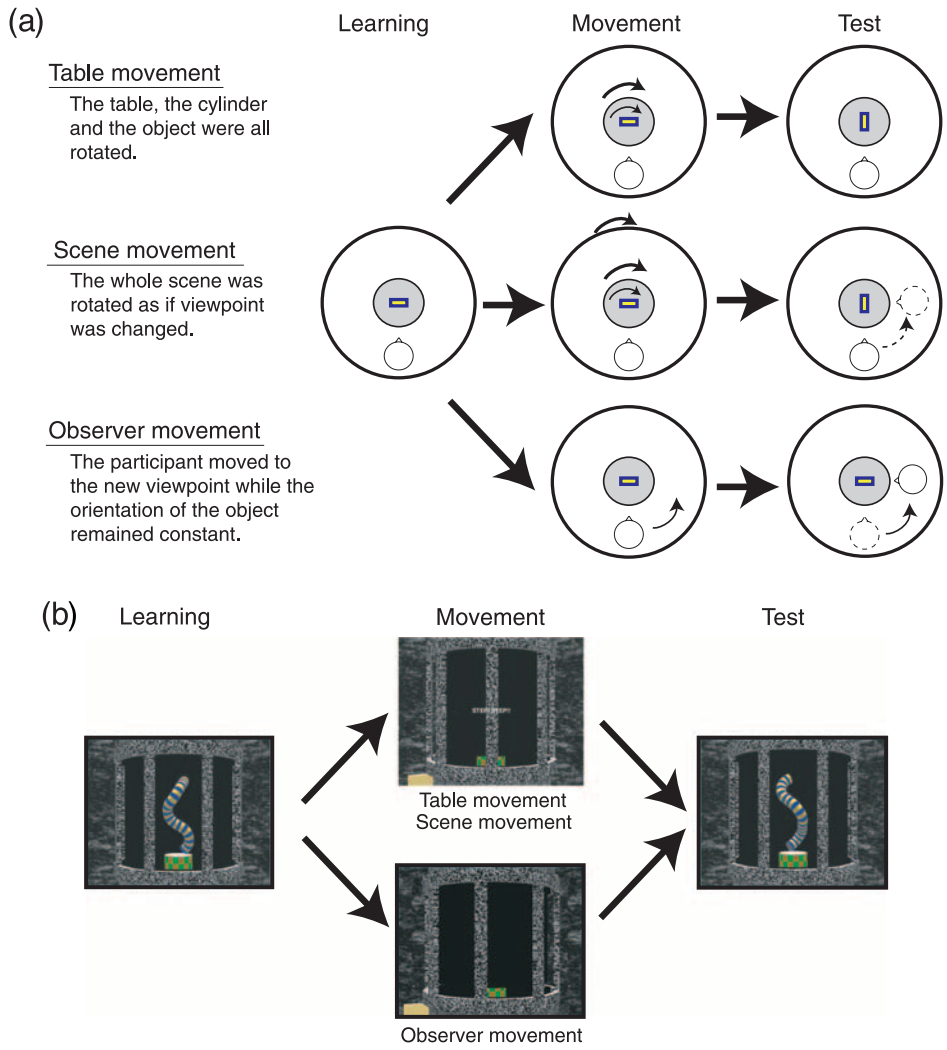

Figure 2. Experimental procedure (a) and examples of a sequence of scenes the HMD displays (b) for Experiment 1 including learning, movement, and test phases.

viewing windows (7 cm × 16 cm) which were located at every 45°. A physical table, whose size and position matched the virtual table, was added to the physical experimental room in order to increase immersion and allow participants to move around the table easily and safely. Twenty novel objects were created by smoothing the edges of Shepard–Metzler's objects using 3ds MAX software (see Figure 2). Ten of them were used for Experiment 1 and the others were used for Experiment 2. The main axis of each object was defined as a line connecting the two farthest vertices of the object. Each object was presented in such a way that the main axis was parallel with the gravity axis. To facilitate shape perception, these objects were textured with blue–yellow stripes. The maximum size of the objects was 14 cm height and 4 cm wide. A light source was centered directly above the objects to provide shape-from-shading cues. All objects were put on a red pedestal with a radius of 2.5 cm inside the cylinder.

Experiment 1

Methods

Design and procedure

We used a 3 (movement condition: observer, whole scene, and table) × 2 (angular disparities: ±45° and ±90°) within-participant experimental design (see Figure 2). Additionally, we included a baseline condition with 0° of angular disparity without participant's viewpoint change. Left and right rotations were randomized to avoid predictability, but not separately analyzed. As described in the Introduction, we compared three movement conditions with each other. In the **observer movement condition**, viewpoint was changed by actual observer's movement around the table while the orientation of the object inside remained unchanged. In the **scene movement condition**, simulated viewpoint change was presented without actual

observer's movement or self-motion perception. In the **table movement condition**, the simulated table, a cylinder with viewing windows and an object on the table was rotated synchronously while the visual background remained stationary. Participant performed sequential matching tasks consisting of three phases: learning, movement, and test phase.

Learning phase: In the learning phase, participants were standing in front of the closed initial window and pressed the "start" button when ready to open it for 2.0 s to reveal a view onto the to-be-learned object. **Movement phase:** On closing the window, a small cue was presented for 1 s under the upcoming viewing window. The participant moved to the indicated window in the observer movement condition, while the window itself moved to the participant's position in the scene movement and table movement conditions. This phase was called "movement phase" and its duration was 2.5 s. A 2.5 s sound clip was presented during the movement phase in order to allow participant to time their movement effectively. To control the effect of movement itself, the participants in the baseline and the table movement conditions made two steps in place in front of the initial window during the movement phase. In these conditions, the text "Step! Step!" was presented on the screen. **Test Phase:** In the test phase, the window was opened again and participants used designated gamepad buttons to judge whether the test object was identical with the initial object or not. Subsequently, participants provided a rating of how sure they were about their previous judgment by adjusting the gamepad slider (which controlled a visually presented slider). The window was not opened unless the participant stood on the correct position and looked in the correct direction (tolerance range: ±5°). Participants saw the test object through the open window until they made a judgment. Thus, the object only appeared in the learning and test phases, separated by either 0, 45, or 90 deg, whereas the rest of the scene remained visible throughout the experiment.

There were 4 sessions, consisting of 78 trials each (20 for each of 3 movement conditions and 18 for the baseline condition). In a session, each of three movement conditions was blocked, including 6 trials for the baseline condition, while the angular disparity was randomized. The order of conditions was randomized between sessions. The movement conditions and angular disparities were randomized within each session. The initial orientation of each object was varied across trials. For half of the trials, the initial object was replaced with one of the different objects (distractor trials), while the same initial object was presented for the other half of the trials (target trials). Before the experimental trials, participants performed two practice sessions of 40 trials (one session with performance feedback and the other without feedback) to become familiar with the setup and procedure. Three hours were needed to complete all sessions. The dependent variables were reaction times between the presentation of the test object and the participants' response, error rates, and confidence ratings. Participants were instructed to make a response as quickly and as accurately as possible.

Results

For each participant and condition, we computed median reaction times and confidence ratings for correct responses for target trials, as well as the error rates for target trials. Figure 3 shows the mean values of those measurands across participants as a function of angular disparity. The data was collapsed across directions of perspective change. Because the baseline condition was conducted in each block of movement condition, it was calculated separately for every movement condition.

A two-way within-participants ANOVA (3 movement conditions × 2 angular disparities) was conducted for each measure. No main effect of movement was observed for any measure [error rate: $F(2, 22) = 0.75$, $p = 0.48$; reaction time: $F(2, 22) = 0.09$, $p = 0.91$; confidence rating: $F(2, 22) = 1.21$, $p = 0.32$]. No interaction effect was observed, either [error rate: $F(2, 22) = 0.39$, $p = 0.68$; reaction time: $F(2, 22) = 2.04$, $p = 0.15$; confidence rating: $F(2, 22) = 0.92$, $p = 0.42$]. A main (or marginal) effect of angular disparity was observed, though [error rate: $F(1, 11) = 3.51$, $p = 0.09$; reaction time: $F(1, 11) = 6.96$, $p = 0.02$; confidence rating: $F(1, 11) = 6.31$, $p = 0.03$]. Thus, this experiment unexpectedly showed no advantage for the observer movement condition at all.

Discussion

The observed lack of any advantage of the observer movement over the other conditions in object recognition might be caused by a variety of factors such as differences in the setup, stimuli, and/or procedure used, as compared to previous studies by Simons et al. (2002) and Zhao et al. (2007). It is, for example, conceivable that object recognition performance might be dependent on the complexity of the objects or on the level of realism and presence for a given virtual environment. In the scene recognition literature, Wang et al. (2006) showed that the number of objects to be automatically updated according to observer's movement could influence object array recognition performance, suggesting that there could be a capacity limitation for the advantage of the observer movement condition. Furthermore, Wang (2004) suggests that there was a possibility that an automatic spatial updating process (which was a hypothesized process responsible for the advantage of the observer movement condition (Simons et al., 2002; Zhao et al., 2007)) might not be activated in a fictitious environment such as a world created by using virtual reality technology. In Wang

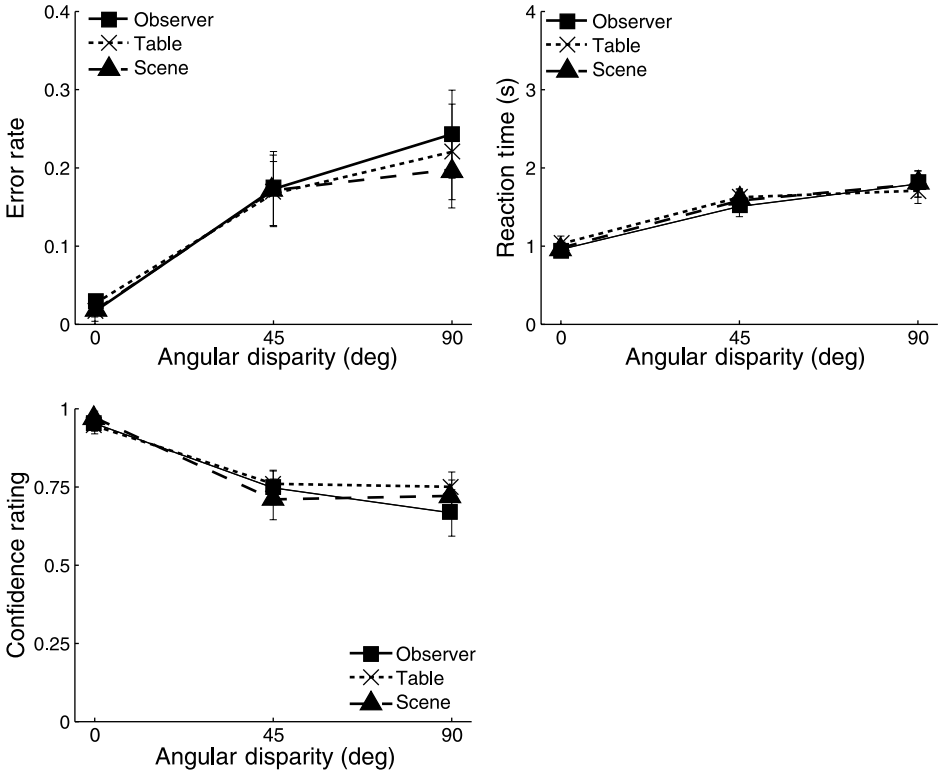

Figure 3. Average reaction time, error rate, and confidence rating as a function of angular disparity in Experiment 1. Note that unexpectedly the movement type did not show any significant effects or interactions for any of the three dependent measures.

(2004), she investigated the difference in the automatic spatial updating process between real and imagined environments. Participants were first required to remember the locations of several objects in a real room and imagine them in their kitchen. Then, they were asked to turn to face one of the objects either in the real room or imagined kitchen and to point to a target located in either the real room or imagined kitchen. The results showed that the objects were automatically updated in the real room even when the participants turned to the objects in the imagined room, while the objects were not updated in the imagined kitchen when the participants turned to the objects in the real room. The finding suggests that the mechanism for automatic spatial updating can depend on the nature of an experimental environment.

There is another major difference in the available information in the movement conditions of Experiment 1 and previous object recognition studies (e.g., Simons et al., 2002; Zhao et al., 2007) that might have contributed to the observed lack of observer movement benefit: Both the scene movement and table movement conditions in Experiment 1 included dynamic visual information about the to-be expected view of the target object. That is, participants in Experiment 1 always saw the scene or table moving to a new orientation, and were thus provided with dynamic visual information (e.g., optic flow) about the angular disparity, whereas the object movement condition in previous studies (e.g., Simons et al., 2002; Zhao et al., 2007) did not include any such dynamic visual information. Note that previous studies did not disambiguate between the influence of dynamic visual and non-visual cues on object recognition from novel views, as the observer movement condition always included both dynamic visual and non-visual (e.g., biomechanical) information about the viewpoint change in conjunction, whereas the object movement condition included neither dynamic visual or non-visual cues about the viewpoint change. To investigate if the dynamic visual information available in all movement conditions of Experiment 1 might have contributed to the unexpected lack of observer-motion benefit, the second experiment compared the observer movement condition (which includes both dynamic visual and non-visual information about the viewpoint change) with an object movement condition similar to previous studies that did not include any dynamic visual or non-visual information about the angular disparity. If we could replicate the observer movement benefit over object movement and the object movement condition would turn out to be worse than the scene and table movement conditions of Experiment 1 (which both included dynamic visual information), then this would indicate the importance of dynamic visual cues, which were present in all but the object movement condition.

Experiment 2

In this experiment, we investigated whether we could replicate the observer movement benefit in object recognition if all dynamic visual information about the viewpoint change were eliminated in an object movement condition. In order to reduce the experimental time per participant, Experiment 2 only used an observer movement and object movement condition and did not repeat the scene movement and table movement conditions of Experiment 1, but instead closely replicated the procedure of Experiment 1 to allow for direct comparison.

While there was dynamic visual information available about the viewpoint change in all conditions of Experiment 1, such dynamic visual information was always coupled with the dynamic non-visual (body motion) information in previous studies like Simons et al. (2002) and Zhao et al. (2007). Comparing the results of Experiments 1 and 2 will allow us to disambiguate the influence of dynamic visual and non-visual information in object recognition from novel views. In particular, if performance in the object movement condition of Experiment 2 should decrease compared to all the other conditions of Experiments 1 and 2 (which all included dynamic visual information about the view change), this would suggest that not only physical motion, but also dynamic visual information can significantly affect object recognition from novel views.

Methods

Design and procedure

We used a 2 (movements: observer and object; randomized) × 2 (angular disparities: ±45° and ±90°; randomized) within-participant experimental design (Figure 4). Additionally, 0° of angular disparity without observer's viewpoint change was also tested as a baseline condition. The learning and the test phases were the same as in Experiment 1, while there were small differences in the movement phase as described below. To indicate whether the current trial was an observer or object motion trial, green or orange arrow(s), respectively, appeared on the screen after the learning phase and throughout the movement phase. For the observer movement condition, green arrow(s) always pointed to the counter-clockwise direction and indicated "go to the next viewing window", which was positioned in that direction. For the object movement condition, orange arrow(s) always pointed to the clockwise direction and indicated that the learned object will be rotating in that direction. The number of arrows indicated the angular disparity for a given trial: One and two arrows indicated 45° and 90° of angular disparity, respectively, while no arrow was presented at 0° of angular disparity (baseline condition). Adding arrows was used to provide participants with explicit information about the angular disparity in the object movement phase, where participants cannot use the biomechanical and visual information available in the observer motion condition. Hence, participants in the object and observer motion condition received similar explicit information about the angular disparity (0°, 45°, or 90°), thus reducing potential confounds. To control for effects of movement itself, participants in the baseline and the object movement condition made two steps in place in front of the initial window. To remind participants of what to do in these conditions, a text "Step! Step!" was presented during the movement phase. Participants saw the test object through the open window until they made a judgment. Thus, the object only appeared in the learning and test phases, separated by either 0, 45, or 90 deg, whereas the rest of

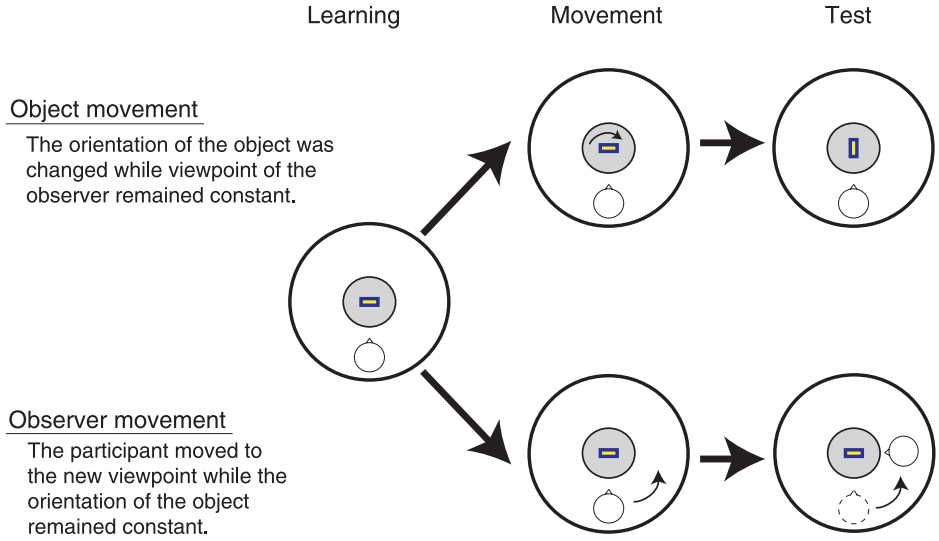

Figure 4. Experimental procedure for Experiment 2.

the scene remained visible throughout the experiment. There were 5 sessions, consisting of 40 trials each. The movement conditions and angular disparities were randomized within each session. Before the experimental sessions, participants performed a practice session of 40 trials without feedback to be familiar with the procedure. The other procedures were the same as those in Experiment 1. Two hours were needed to complete all sessions.

Results

For each participant and condition, we computed the median reaction times and confidence ratings for correct responses for target trials, as well as the error rates for target trials. Figure 5 shows the average values of error rates, reaction times and confidence ratings across participants as a function of angular disparity. Note that for Experiment 2, the data for 0° of angular disparity (i.e., the baseline condition) is identical between the two movement conditions, different from the baseline condition of Experiment 1. A two-way within-participants ANOVA (2 movements × 2 angular disparities) was conducted for each measure. A significant main effect of movement was observed for all measures—error rate, $F(1, 11) = 7.37$, $p = 0.02$; reaction time, $F(1, 11) = 4.72$, $p = 0.05$; confidence rating, $F(1, 11) = 6.48$, $p = 0.03$. That is, for all measures, the performance for the observer movement condition was better than that for the object movement: lower error rates and reaction times, and higher confidence ratings for the observer movement condition. A marginal effect of angular disparity was observed for reaction time, $F(1, 11) = 4.05$, $p = 0.07$, but not for error rate, $F(1, 11) = 0.04$, $p = 0.85$ or confidence ratings, $F(1, 11) = 1.57$, $p = 0.24$. No interaction effect was observed for any measure, $Fs(1, 11) < 0.82$, $ps > 0.38$. These results indicate an advantage of the observer movement condition over the object movement condition for both 45° and 90° angular disparity. To compare the performance for the baseline condition with the movement conditions, we conducted a one-way ANOVA (factor: angular disparity) for each movement condition for each measure. The analysis revealed that the performance for 0° of angular disparity was significantly better than that for 45° and 90° for both movement conditions and for all measures, $Fs(2, 22) > 7.81$, $ps < 0.005$. Finally, in order to compare performance between Experiments 1 and 2 and thus to disentangle the influence of dynamic visual and non-visual information, a two-way ANOVA (5 movement

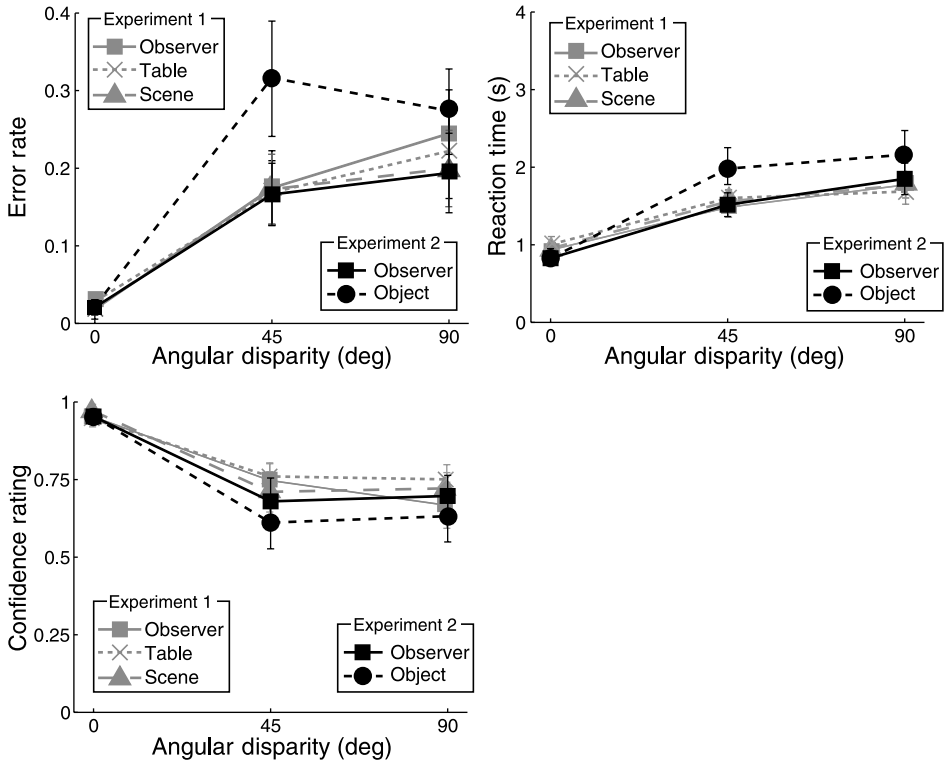

Figure 5. Average reaction time, error rate, and confidence rating as a function of angular disparity in Experiment 2 (solid black). For easier comparability, we included the data from Experiment 1 (grayshaded). Note that for Experiment 2, the data for 0° of angular disparity (i.e., the baseline condition) is identical between the two movement conditions and that the object movement condition showed significant performance decrements compared to the observer movement condition in all three dependent measures ($p < 0.05$), but no interactions with angular disparity (45 vs. 90 deg).

conditions × 2 angles) and its post hoc analysis (Tukey's HSD, $p < 0.05$) were conducted for each measure. The object movement condition showed less accurate performance and lower confidence ratings than any of the other conditions—error rate, $F(4, 44) = 2.88$, $p = 0.03$; reaction time, $F(4, 44) = 2.01$, $p = 0.10$; confidence ratings, $F(4, 44) = 2.87$, $p = 0.03$. No interaction was observed for any measure, $Fs(4, 44) < 1.07$, $ps > 0.38$.

Discussion

The main finding of this experiment was that object recognition performance was more accurate and quicker when new views resulted from observers' own movements than when they resulted from object movements. These results are consistent with the previous studies (Simons et al., 2002; Zhao et al., 2007), replicating the advantage of the observer movement over the object movement condition for our experimental setup and stimuli, thus validating our procedure. However, it is notable that the current results were different from those of the previous studies in two ways. First, Simons et al. (2002) observed comparable recognition performance between the 0-deg and 40-deg rotation conditions in the movement condition, which we did not. Second, there was no interaction effect between type of movement and angular disparity in the current study, meaning that the observer movement facilitated object recognition across angular disparities (45° and 90°), while Zhao et al. (2007) observed the facilitation only for 50°, but not for 90°. As suggested in the introduction of this experiment, these differences might be caused by differences in the complexity of the objects or on the level of realism and presence for a given virtual environment (Wang, 2004; Wang et al., 2006). Although there were small differences from the previous studies as described above, this experiment showed that our virtual reality environment was appropriate for observing the advantage of the observer movement condition in object recognition.

Combining the results from Experiments 1 and 2 allows us to disentangle the influence of dynamic visual and non-visual motion cues for object recognition from novel views: Performance in the observer movement conditions in Experiments 1 and 2 was virtually identical, and matched performance in the scene and table movement condition which both included dynamic visual information about the disparity change, but no non-visual (physical motion) cues. All of those conditions (which included dynamic visual information) outperformed the object movement condition of Experiment 2. Together, the data can be interpreted as evidence that the lack of benefit of the observer movement over the scene and table movement conditions in Experiment 1 was caused not by decreasing the performance for the observer movement condition, but instead by improving the performance for the scene and table movement conditions with respect to the object movement condition (which was the only condition that did not include dynamic visual information). Interestingly, though, adding non-visual motion cues in the observer movement condition of Experiment 1 did not provide additional object recognition benefits compared to the scene and table movement condition, which provided dynamic visual cues but did not include observer motion (non-visual cues). In summary, this means that dynamic visual information about the amount of perspective change, regardless of whether the change is caused by orientation change of an object or viewpoint change of an observer, can be one of the crucial factors responsible for the advantage of the observer movement condition over the object movement condition.

General discussion

Previous studies showed that object recognition performance was improved when perspective change resulted from observer movement instead of object movement (Simons et al., 2002; Zhao et al., 2007). Here, our goal was to investigate what specific factors actually cause this benefit of observer movement over object movement. That is, we tested which kind of information related to observer's movement was crucial for the phenomenon. In previous studies, information about the magnitude of perspective change were much richer in the observer movement condition than in the object movement condition (Simons et al., 2002; Zhao et al., 2007): participants could use visual, vestibular and proprioceptive signals as well as efference copies of motor commands to compute the magnitude of the perspective change, and each of these cues might have contributed to the observed advantage of observer movements over object movements. Therefore, there is a possibility that the observed benefit of observer movements over object movements could be sufficiently explained by differences in the amount of information available to the participant, even if we do not assume any specialized mechanisms such as automatic spatial updating.

While previous studies have typically attributed the observer motion benefit to biomechanical and vestibular (i.e., non-visual) motion cues, we questioned this prevailing opinion by testing to what degree dynamic *visual* information might contribute to the observer movement benefit in object recognition. Apart from the classic object movement and observer movement condition, we also included two ways of providing participants with dynamic visual information without any non-visual motion cues: In the scene movement condition, the whole visual scene was rotated without any actual observer's movement or any (illusory) self-motion perception. That is, the visual cues are very similar to the observer movement condition, but without any actual observer motion. In the table

movement condition, only the virtual table and the cylinder on the table rotated while the visual background and the participant remained stationary. Comparing these different conditions allowed us to investigate the relative contribution of dynamic visual cues and physical motion cues independently.

The data showed a clear benefit of observer movements over mere object movements (Experiment 2), replicating previous results (Simons et al., 2002; Zhao et al., 2007). Unexpectedly, however, both the scene movement and table movement condition showed a similar benefit over the object movement condition, despite the lack of any non-visual (physical) motion cues, and performance levels equaled those of the observer movement condition (Experiment 1). Thus, at least for the current experimental paradigm and setup, all of the observer movement benefit might be explained by the dynamic visual information provided, regardless of whether the perspective change was caused by an orientation change of the table containing the object or by a simulated viewpoint change of an observer. That is, the additional physical motion cues provided in the observer motion condition did not provide any additional performance benefit over the dynamic visual information alone.

It is critical to note that the table movement condition did not include any real or implied self-motion information (as the visual background remained stationary), while the scene and observer movement conditions did (even though none of the participants experienced any illusory self-motion perception (vection) in the scene movement condition). This means that the current data can be fully explained without any need to assume any automatic spatial updating mechanisms as hypothesized by previous studies, and further research is needed to investigate what—if any—contribution automatic spatial updating has for object or scene recognition from novel viewpoints. Furthermore, active control of observer movement can be excluded as a potential contributing factor in this study, as both the table and scene movement conditions were totally passive yet resulted in performance levels equaling the active (observer movement) condition. This is in agreement with results by Wang and Simons (1999) that found no benefit of active over passive motions for change detection in object arrays from novel viewpoints.

There are (at least) two possibilities for how dynamic visual information can improve object recognition. The first is that a view-transformation system such as "mental rotation" (Shepard & Cooper, 1982; Shepard & Metzler, 1971) underlies all perspective changes, and that dynamic visual information which indicated how much a view must be changed assisted this mental transformation mechanism to match the second (or test) retinal image with the initial (or learned) image. There are several studies showing that object recognition performance can indeed be improved by additional information about the difference between the learned view and the test view (Christou et al., 2003; Cooper & Shepard, 1973). For example, Cooper and Shepard (1973) presented a 2D shape and then gave their participants an indication of the orientation to which the shape might be rotated with various time intervals before the presentation of the test stimuli. Their study revealed that handedness judgments (discrimination between original and mirror reflected characters) were improved with increasing time to pre-process the orientation indicator. For pre-processing times of 1 s, performance was virtually view-independent. Christou et al. (2003) used a desktop virtual room to investigate the effect of environment context (implicit indication of the amount of perspective change) on object recognition. They showed that a well-learned room background could serve as an implicit indication of the observer's new viewing position which helped to recognize the object. Furthermore, performance was only improved when the object was presented in the original (learned) environment, but not when the object was presented in the environment with uniform colored background or with wrongly displaced background. Note, however, that object recognition performance remained view-dependent.

Although these studies indicate that the information about the upcoming perspective can facilitate object recognition performance, there are other studies that showed no such benefit, including Experiment 2 of this study. For example, Cooper and Shepard (1973) showed that up to 1 s was needed to make full usage of the advance information of the upcoming perspective, and 100 ms pre-processing times yielded only minimal benefit. In Simons et al. (2002), as mentioned in the Introduction, a visual background that included unique landmarks did not facilitate object recognition from novel viewpoints. These studies indicate that the indication may be useless unless it is presented well before the presentation of test objects (Cooper & Shepard, 1973) and unless it is visually salient enough (Simons et al., 2002). As the arrows used in Experiment 2 of our study to indicate the amount and direction of perspective change were quite large and colored, and were presented for 2.5 s, we would argue that they were probably both salient enough and presented for a sufficient amount of time (1 s proved sufficient in the study by Cooper & Shepard, 1973). Nevertheless, the arrows did not compensate for the disadvantage of object movements for object recognition. However, providing dynamic visual information (e.g., in the table movement condition) clearly improved performance. Hence, just providing static advance information about the to-be-expected perspective switch does not seem to be sufficient in our case, and mental perspective switches/transformations seem to be better facilitated by dynamic visual information. It has recently been reported that the sequence of views of a three-dimensional object (i.e., dynamic visual information provided by the movement of the to-be-learned object per se) can improve recognition/identification of the object (e.g., Balas & Sinha, 2009; Friedman, Vuong, & Spetch, 2009; Liu, 2007; Mitsumatsu & Yokosawa, 2003; Stone, 1998, 1999; Vuong, Friedman, & Plante, 2009; Vuong &

Tarr, 2004). Several of these studies suggested that motion information could allow observers to predict upcoming views (e.g., Friedman et al., 2009; Mitsumatsu & Yokosawa, 2003; Vuong & Tarr, 2004). Thus, we propose that the dynamic visual information might be easier to use for mental transformation and/or to predict upcoming views than just the static arrows, because it can somehow "prime" or more intuitively "show" how to transform the representation of the object the participants learned.

The second possibility explaining how dynamic visual information can improve object recognition is that the dynamic visual information made it easier for the participants to access the 3D (like geon, Biederman, 1987; Hummel & Biederman, 1992) or 2.5D (Marr, 1982) representations of objects in the brain. In almost all studies in the object recognition literature, objects were statically presented on computer screens and the participant's head was more or less stationary in front of the display. In these experimental situations, it might be hard to construct full 3D or 2.5D representations of objects in the brain even when various depth cues (shading, texture and stereo) were provided, as was done in Edelman and Bülthoff (1992). One possibility is that the dynamic visual information improved figure-ground segmentation and thus made the boundary between the foreground and the background more salient (see Todd, 1995, for a review of structure-from-motion), which, in turn, might make it easier to construct those representations. Considering the increase in reaction times and error rates between the 0° and 45° (and 90°) condition even with the dynamic visual information, however, participants might not have been able to construct or access the full 3D or 2.5D mental representations of a given object and/or might have used additional cues to facilitate object recognition in the 0° condition, such as the 2D shape of the occluding boundary, or the particular 2D pattern of coloration for the 0-deg rotation condition.

The previous studies suggest that a spatial updating mechanism is the most promising explanation underlying the advantage of observer movement over object movement conditions. However, this study revealed that dynamic visual information, but not movement of the whole visual scene as evoked during self-motion, can be a crucial factor for it. Because the observer movement condition in the previous studies also included dynamic visual information, most of the advantage of the observer movement condition seems to be sufficiently explained by this factor. However, different from Simons et al. (2002), we did not observe a complete elimination of view-dependency for object recognition in the observer movement condition in our study. This may be because the level of realism and presence in our virtual reality environment was not as high for real environments, and cannot activate our powerful ability to update the representation of an object in the brain based on self-motion information, as suggested by Wang (2004). In other words, the advantage of the observer movement condition observed here might be different from that in Simons et al. (2002). To elucidate this point, further experiments in a real environment are required. Alternatively, differences in the objects used might also have contributed to the lack of view independency in our study: While Simons et al. (2002) used different rectangular arrangements of wooden squares, we used smoothly curved objects without any clear distinguishing features apart from their geometric shape, which arguably made it harder to use abstract/cognitive strategies (like counting blocks).

Conclusion

In conclusion, the current results question the prevailing opinion that active control and physical motion cues are essential for effective object/scene recognition, spatial orientation, and navigation in VR, and suggest that, at least for some tasks, dynamic visual information might be sufficient—even when presented through a low-cost HMD with a rather limited FOV and resolution. In particular, our results question the typically claimed importance of physical motion-based automatic spatial updating of the observer position in object recognition from novel viewpoints. Note that our results were obtained using an experimental paradigm that was until now predominately thought to be a perfect exemplar of a case where automatic spatial updating based on physical motion cues was essential. These results contribute to the accumulating evidence that naturalistic visual cues in VR can under certain conditions suffice for enabling effective spatial orientation from novel viewpoints (Riecke, Cunningham, & Bülthoff, 2006; Riecke, von der Heyde, & Bülthoff, 2005).

Acknowledgments

Preliminary results of this study were presented at European Conference on Visual Perception held in Arezzo, Italy, in August 2007.

This research was supported in part by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Specially Promoted Research (no.19001004), and the Max Planck Society.

Commercial relationships: none.
Corresponding author: Wataru Teramoto.
Email: teraw@ais.riec.tohoku.ac.jp.
Address: Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira Aoba-ku Sendai, 980-8577, Japan.

References

Balas, B., & Sinha, P. (2009). A speed-dependent inversion effect in dynamic object matching. *Journal of Vision, 9*(2):16, 1–13, http://www.journalofvision.org/content/9/2/16, doi:10.1167/9.2.16. [PubMed] [Article]

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Bülthoff, H. H., Edelman, S., & Tarr, M. J. (1995). How are 3-dimensional objects represented in the brain? *Cerebral Cortex, 5,* 247–260.

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

Christou, C. G., Tjan B. S., & Bülthoff, H. H. (2003). Extrinsic cues aid shape recognition from novel viewpoints. *Journal of Vision, 3*(3):1, 183–198, http://www.journalofvision.org/content/3/3/1, doi:10:1167/3.3.1. [PubMed] [Article]

Cooper, L. A., & Shepard, R. N. (1973). Time required to prepare for a rotated stimulus. *Memory and Cognition, 1,* 246–250.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32,* 2385–2400.

Friedman, A., Vuong, Q. C., & Spetch, M. L. (2009). View combination in moving objects: The role of motion in discriminating between novel views of similar and distinctive objects by humans and pigeons. *Vision Research, 49,* 594–607.

Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology, 9,* 1315–1318.

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

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

James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Boddour, R., & Goodale, M. A. (2002). "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, and Computers, 34,* 383–390.

Liu, T. (2007). Learning sequence of views of three-dimensional objects: The effect of temporal coherence on object memory. *Perception, 36,* 1320–1333.

Marr, D. (1982). *Vision*. New York: Freeman.

Mitsumatsu, H., & Yokosawa, K. (2003). Efficient extrapolation of the view with a dynamic and predictive stimulus. *Perception, 32,* 969–983.

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology, 58,* 75–96.

Riecke, B. E., Cunningham, D. W., & Bülthoff, H. H. (2006). Spatial updating in virtual reality: The sufficiency of visual information. *Psychological Research, 71,* 298–313.

Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2005). Visual cues can be sufficient for triggering automatic, reflex-like spatial updating. *ACM Transactions on Applied Perception, 2,* 183–215.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

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

Simons, D. J., Wang, R. X. F., & Roddenberry, D. (2002). Object recognition is mediated by extraretinal information. *Perception and Psychophysics, 64,* 521–530.

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research, 38,* 947–951.

Stone, J. V. (1999). Object recognition: View-specificity and motion-specificity. *Vision Research, 39,* 4032–4044.

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

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience, 1,* 275–277.

Todd, J. T. (1995). The visual perception of three-dimensional structure from motion. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 201–226). New York: Academic Press.

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

Vuong, Q. C., Friedman, A., & Plante, C. (2009). Modulation of viewpoint effects in object recognition by shape and motion cues. *Perception, 38,* 1628–1648.

Vuong, Q. C., & Tarr, M. J. (2004). Rotation direction affects object recognition. *Vision Research, 44,* 1717–1730.

Wang, R. X. F. (2004). Between reality and imagination: When is spatial updating automatic? *Perception & Psychophysics, 66,* 68–76.

Wang, R. X. F., Crowell, J. A., Simons, D. J., Irwin, D. E., Kramer, A. F., Ambinder, M. S., et al. (2006). Spatial updating relies on an egocentric representation of space: Effects of the number of objects. *Psychonomic Bulletin and Review, 13,* 281–286.

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

Zhao, M. T., Zhou, G. M., Mou, W. M., & Hayward, W. G. (2007). Spatial updating during locomotion does not eliminate viewpoint-dependent visual object processing. *Visual Cognition, 15,* 402–419.