

# Body and head tilt reveals multiple frames of reference for spatial attention

Yuhong V. Jiang

Department of Psychology, University of Minnesota,  
Minneapolis, MN, USA



Khena M. Swallow

Department of Psychology, Cornell University,  
Ithaca, NY, USA



**Most modern theories of spatial attention suggest that it is based on a maplike representation that prioritizes information in some spatial locations over others. However, movement through space changes the relationship between what is “out there” and a person’s viewpoint. Does spatial attention move with the viewer, or does it stay in environmental locations? Several recent psychophysical and neuroscience studies have attempted to address this question by probing attention following saccadic eye movements. The alignment of the head and body to the external environment in these studies, however, makes it impossible to determine whether attention is based on the viewer’s location in space or on the external environment. The current study therefore introduces a head and/or body tilt through the vertical plane to dissociate viewer-centered from environment-centered representations. Participants first acquired a long-lasting attentional bias to a region of the search display that was likely to contain a target. They then tilted their head or body, and the location of the spatial bias was evaluated. The results suggest that attention has both a viewer-centered component that rotates with the viewer’s head and an environment-centered component that is tied to environmental locations.**

Whereas some studies have revealed lingering attentional effects at the same environmental location (“spatiotopic” representation; Mathôt & Theeuwes, 2010; Maylor & Hockey, 1985; Pertzov, Avidan, & Zohary, 2011; Pertzov, Zohary, & Avidan, 2010; Posner & Cohen, 1984), others have found effects at the same retinal location (“retinotopic” representation; e.g., Abrams & Pratt, 2000; Golomb, Chun, & Mazer, 2008; Golomb & Kanwisher, 2012; Harrison, Mattingley, & Remington, 2012). Yet, regardless of the outcome of such studies, their scope is limited in at least two ways. First, only the eyes moved. The head and body of the participant always aligned with the environment, making it impossible to dissociate environment-centered from head- or body-centered representations. Second, these studies tested transient forms of spatial attention whose effects peak and disappear within several hundred milliseconds. Yet many forms of attention, especially those acquired through visual statistical learning, persist over minutes or even days (Chun & Jiang, 2003; Jiang, Swallow, Rosenbaum, & Herzig, 2013). Because changes in viewpoint are inevitable over such a long time, the reference frame of these durable forms of attention is of great functional significance.

## Introduction

Recent psychophysical and neurophysiological studies have probed the coordinate systems in which an attended spatial location is coded (for reviews, see Cavanagh, Hunt, Afraz, & Rolfs, 2010; Wurtz, 2008). In these studies, participants were cued to attend to one location. An eye movement was then made, and spatial attention was probed either at the same retinotopic location or at the same environmental location as the cue. The data from these studies have been mixed:

This study examines the coordinate systems of durable attention by introducing body and/or head tilt. Participants conduct visual search for a T target among L distractors presented on an upright computer monitor. Unbeknownst to participants, across multiple trials, the T target is more often found in a high-probability, “rich” quadrant than in the other “sparse” quadrants. Previous research has shown that participants rapidly acquire a spatial bias toward the rich quadrant (Druker & Anderson, 2010; Geng & Behrmann, 2002; Umemoto, Scolari, Vogel, & Awh, 2010). In addition, this spatial bias persists for several hundred trials of extinction training (Jiang, Swallow,

Citation: Jiang, Y. V., & Swallow, K. M. (2013). Body and head tilt reveals multiple frames of reference for spatial attention. *Journal of Vision*, 13(13):9, 1–11, <http://www.journalofvision.org/contents/13/13/9>, doi:10.1167/13.13.9.

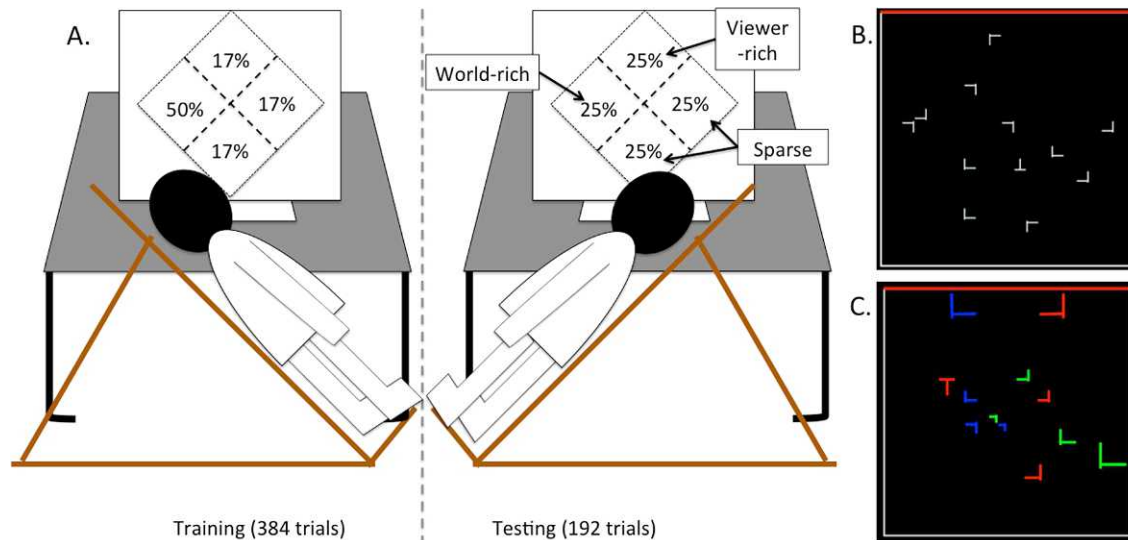


Figure 1. (A) Design and setup of Experiment 1. Participants performed visual search while resting against a stand, tilting their body  $45^\circ$ . Body tilt changed  $90^\circ$  between the training and testing phases. The target's location probability also changed. (B) A sample search display in Experiments 1–3. (C) A sample search display in Experiment 4; item size was adjusted according to the cortical magnification factor.

Rosenbaum, et al., 2013). To probe whether the spatial bias is directed toward the same environmental locations or whether it is viewpoint dependent, participants tilt their body and/or head  $90^\circ$  through the vertical plane. Following the tilt, they complete a testing phase involving the same visual search task but with random (unbiased) target locations. This manipulation produces three types of quadrants on the display: (a) the quadrant on the screen that was rich during training (world-rich quadrant), (b) the quadrant in the participant's visual field that was rich during training (viewer-rich quadrant), and (c) the remaining two quadrants that were sparse during training (Figure 1A). Of interest is where the spatial attention bias that developed during training is located after the change in body and/or head orientation.

One advantage of the design used here is that unlike changes in eye position, changes in body and/or head orientation fully dissociate body- and head-centered reference frames from environment-centered reference frame. Tilting the participant's head or body after the persistent attentional bias to a quadrant has been acquired allows a test of whether the bias is located at the same location relative to the environment or whether it is in the same location relative to the participant's head (or body). Computations that allow attended locations to remain in the same environmental location (e.g., spatiotopic coding or spatial updating) should lead to faster target detection when the target appears in the same place on the monitor as where it was most often found in the past. In contrast, computations that code attended locations relative to the viewer, but without spatial updating, should yield

faster target detection when the target appears in the same part of the viewer's visual field as where it was most often found in the past.

The four experiments in this study differ in a number of dimensions, including the degree of tilt from the vertical axis during training ( $45^\circ$  or  $0^\circ$ ), the alignment of the body and the head (whole body tilt or head tilt only), and whether the display is presented long enough for eye movements to play a role in search (unlimited viewing duration or brief presentation). However, the findings consistently suggest the use of two reference frames in attention: an environment-centered reference frame and a viewer-centered reference frame. These behavioral data provide a basis for future neuroimaging and neurophysiological research on the neural substrates for the two coordinate systems of spatial attention.

## Method

### Participants

Participants in all experiments were students at the University of Minnesota between the ages of 18 and 35 years. All participants were naïve to the purpose of the study, had normal or corrected-to-normal visual acuity, passed a color blindness test, and participated in no more than one experiment. The research adhered to the tenets of the Declaration of Helsinki and was approved by the University of Minnesota's Institutional Review Board. All participants signed a written consent before

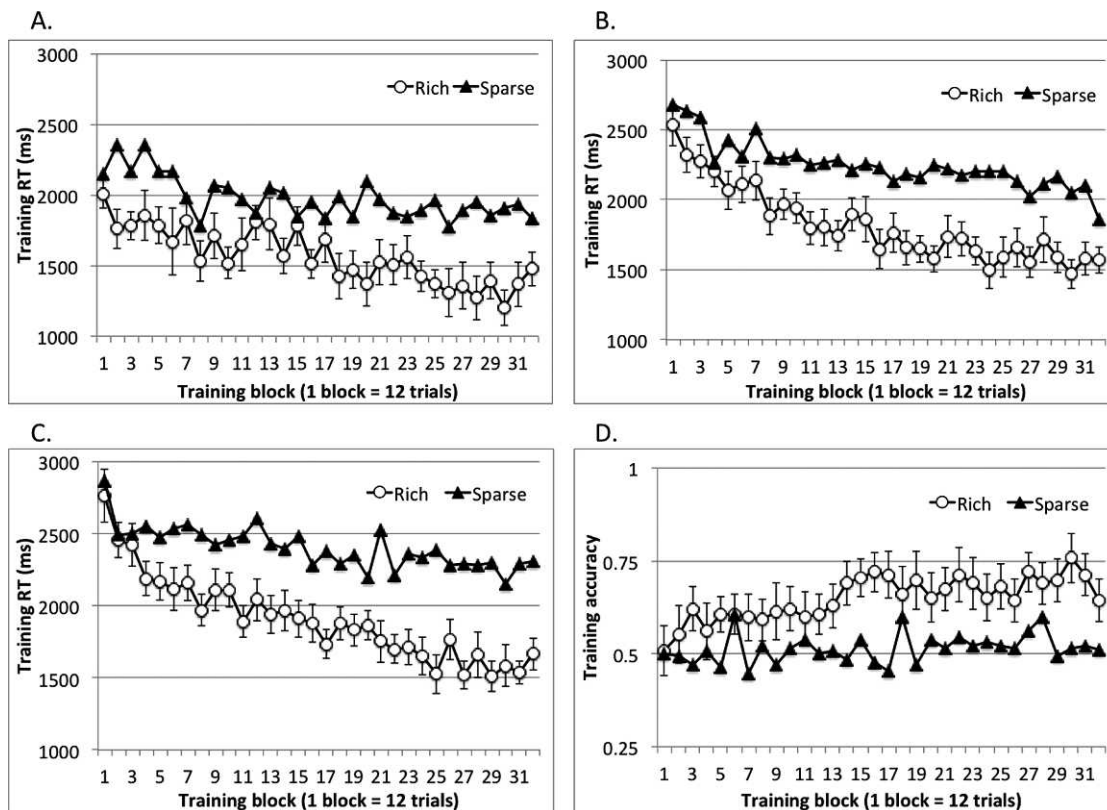


Figure 2. Results from the training phase. (A) Experiment 1's RT. (B) Experiment 2's RT. (C) Experiment 3's RT. (D) Experiment 4's accuracy. Error bars show  $\pm 1$  SE of the difference between the rich and sparse conditions.

the experiment. Participants received \$10/hr or extra course credit for their participation. The number of participants was 16 in Experiment 1, 32 in Experiment 2, 32 in Experiment 3, and 22 in Experiment 4.

## Equipment

Participants were tested individually in a room with normal interior lighting. Stimuli were presented on an upright 17-in. CRT monitor (1024  $\times$  768 pixels, 75-Hz vertical refresh rate). The experiment was programmed using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Participants responded with an optical wireless mouse.

In Experiment 1, participants rested on the slanted surface (6.5-ft. long and 1-ft. wide) of a wooden triangular structure, tilting their body and head exactly 45° from vertical (Figure 1A). Padding raised the participant's head so it was parallel to the participant's body and the surface. Viewing distance was between 107 and 117 cm but varied with the participant's height.

In Experiment 2, participants sat upright in the training phase with a viewing distance of approximately 110 cm. In the testing phase, they rested flat on their side on a long table. Cushions and a small pillow raised the participant such that his or her head was

roughly in the same position in the training and testing phases (approximately 120 cm above the floor). Trials were computer paced for the first 16 participants (1.5 s between the end of one trial and the beginning of the next) and self-initiated for the last 16 participants. Results were pooled because they were unaffected by trial pace.

In Experiment 3, participants tilted their head 45° from vertical, resting their head on the slanted surface of a headrest. Their body was upright in the entire experiment. The headrest was adjusted to align the participants' nose with the center of the monitor. The headrest and monitor were on the same 75-cm-long table, limiting the viewing distance to 38 cm. For the first 16 participants, the headrest had a fixed 45° slope. An experimenter adjusted the participants' head orientation until their head was tilted 45° as it rested against the headrest. Because the facial bones are not completely flat, some participants had to exert extra effort to keep their heads oriented at the desired angle. The next 16 participants were tested with an adjustable headrest, whose angle was adjusted to yield a 45° head tilt without discomfort. Results were pooled because statistical analyses showed no effect of the device on the data.

Experiment 4 used the same device as was used for Experiment 3. Fourteen participants were tested with

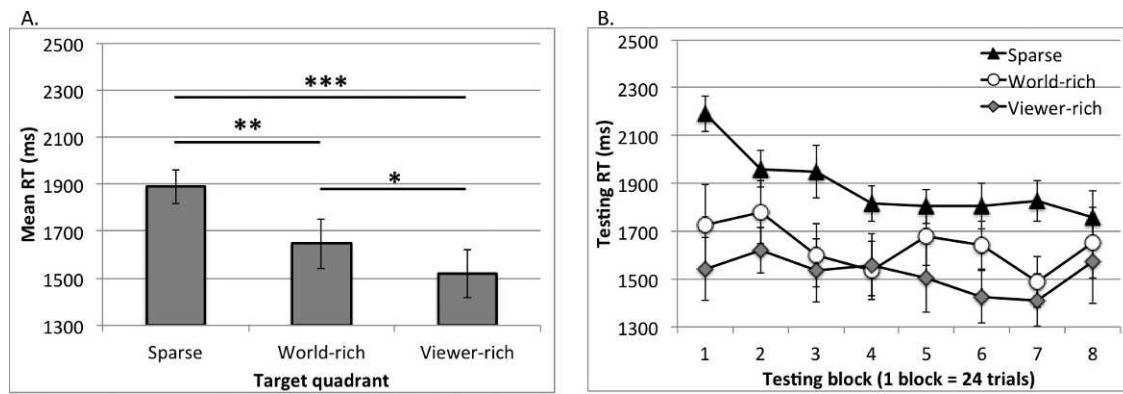


Figure 3. Results from Experiment 1’s testing phase. (A) Mean across all trials. (B) Data divided into eight blocks. Error bars show  $\pm 1$  SE of the mean. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

the 45° fixed-slope headrest, whereas the other 8 were tested with an adjustable-slope headrest.

**Materials**

Participants searched for a T target among 11 L distractors. The orientation of each item was randomly drawn from 0°, 90°, 180°, or 270° to ensure that the items did not change their appearance after the body/head rotation. A white outline square (700 × 700 pixels) framed the search space. One side of the square was red to provide a stable environmental landmark.

In Experiments 1, 2, and 3, the display stayed on the screen until participants responded to the target’s color. Each item subtended 40 × 40 pixels (0.71° × 0.71° at a viewing distance of 110 cm). All items were white but tinted slightly red [RGB: 255 240 240] or green [RGB: 225 255 225]. The background was black. Each display contained 12 items. The locations of the search items were randomly selected from a 10 × 10 invisible matrix.

The axes of the matrix aligned with the participants’ head orientation. This matrix was divided into quadrants (Figure 1B), and search items were distributed equally across quadrants (three items/quadrant).

In Experiment 4, the display lasted approximately 180 ms (the precision was limited by the screen’s refresh rate). The display was not masked. Items had saturated red [255 0 0], green [0 255 0], or blue [0 0 255] color, randomly selected for each item. Item sizes were scaled according to a cortical magnification factor to compensate for decreased visual acuity at greater eccentricities (Carrasco, Evert, Chang, & Katz, 1995). Items were placed on four rings with a radius of 50, 120, 200, or 350 pixels from fixation (2.6°, 6.2°, 10.3°, and 17.9° at a viewing distance of 38 cm). Each ring contained eight equidistant locations. The 12 search items were placed in randomly selected locations among the set of 32 possible locations, with the constraint that there were three items per quadrant (Figure 1C).

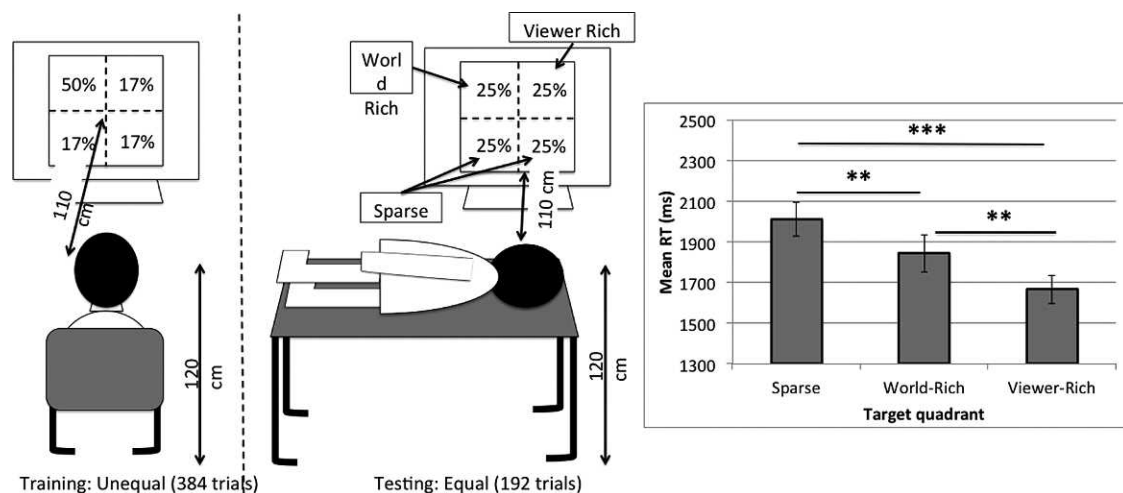


Figure 4. (Left) A schematic illustration of the experimental design used in Experiment 2. (Right) Testing phase results from Experiment 2. Error bars show  $\pm 1$  SE of the mean. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

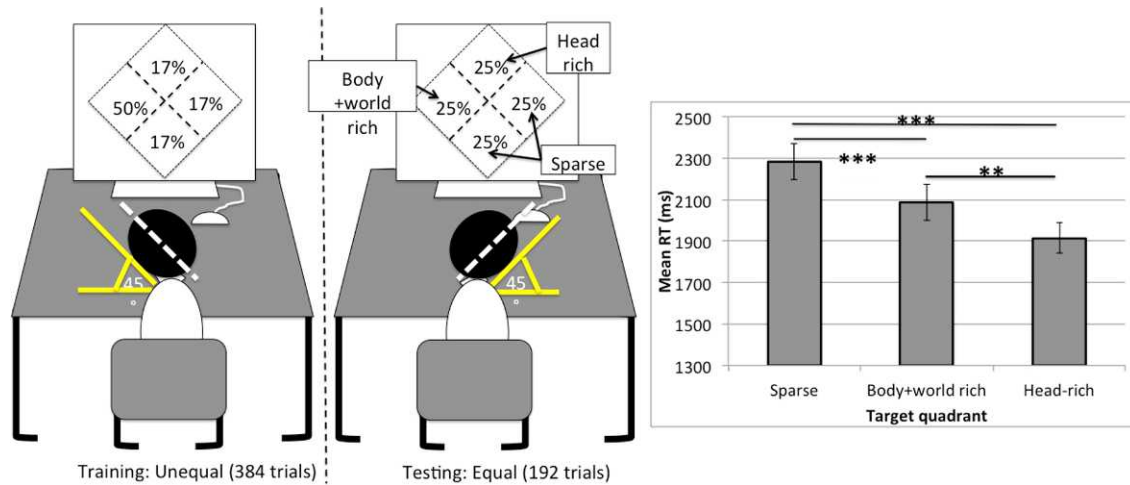


Figure 5. (Left) Experimental setup and design used in Experiments 3 and 4. (Right) Results from the testing phase of Experiment 3. Error bars show  $\pm 1$  SE of the mean.  $**p < 0.01$ ;  $***p < 0.001$ .

## Task and procedure

Each trial started with a white central fixation square ( $10 \times 10$  pixels) whose position changed slightly from trial to trial (by up to 50 pixels from the display center). Participants were asked to fixate on the square. Participants then clicked on the fixation square with the mouse to initiate the trial. The mouse click required eye-hand coordination and brought fixation to the center of the display. After a 200-ms delay, the search display was presented. Participants were asked to find the T and report its color with the corresponding mouse button. The target's color was randomly chosen for each trial, so the response was not associated with any experimental manipulations. Three rising tones (300 ms total) followed a correct response. A buzz (200 ms) and a 2-s blank display followed an incorrect response.

## Design

After 10 practice trials with a randomly located target, participants completed 384 trials of training and 192 trials of testing. During training, the target appeared in a high-probability “rich” quadrant on 50% of the trials and in any one of the low-probability “sparse” quadrants on 16.7% of the trials. Which quadrant was rich was fixed for a given participant and counterbalanced across participants.

Two changes occurred between training and testing. First, the target was now equally likely to appear in each of the four quadrants (25% of the trials). Second, the participants' body and/or head rotated through the vertical plane by  $90^\circ$ . In Experiment 1, participants stepped off the wooden structure that tilted them  $45^\circ$  in one direction. The experimenter then turned the

wooden structure around and participants returned to it, now leaning  $45^\circ$  in the opposite direction. This change produced a  $90^\circ$  rotation of the body. The direction of body rotation (clockwise or counterclockwise) was counterbalanced across participants. In Experiment 2, participants changed from sitting upright to lying down on a table. Half of the participants rested on their left side, and the other half rested on their right side during testing. In Experiments 3 and 4, the headrest was turned around, and participants rested their head  $45^\circ$  in the opposite direction as in training. The direction of head tilt was counterbalanced across participants. The rotation procedure that occurred before the testing phase took several minutes to complete in all experiments. Participants were free to move around during that time.

In none of the experiments did we tell participants where the target was likely to appear at any point. An experimenter stayed in the room to monitor the participants' body and/or head orientation.

At the end of the study, participants completed a recognition test while remaining in the same orientation that they were in during testing. The experimenter first queried informally about whether the participants thought the target was equally likely to appear in any part of the display. Explicit knowledge was then assessed formally by asking participants to click on where they thought the target was most often found.

## Results

### Training phase

We first examined whether a spatial bias emerged in the training phase. In this phase, the target appeared in

a high-probability “rich” quadrant three times more often than in any one of the low-probability “sparse” quadrants. This manipulation yielded a strong spatial bias toward the rich quadrant (Figure 2).

Specifically, in Experiments 1 to 3, in which the items remained visible until a response was made, search response time (RT; excluding incorrect trials and trials with an RT longer than 10 s or shorter than 200 ms) was significantly faster when the target was in the rich quadrant than the sparse quadrants,  $F(1, 15) = 32.49$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.68$ , in Experiment 1;  $F(1, 31) = 152.99$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.83$ , in Experiment 2; and  $F(1, 31) = 120.27$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.80$ , in Experiment 3. The overall RT was longer in Experiment 3 than in Experiments 1 and 2, possibly reflecting the shorter viewing distance and the fact that the display subtended a larger visual angle. In addition, breaking the training phase into 32 blocks, we found a significant interaction between target quadrant and the linear trend of block,  $F(1, 15) = 6.04$ ,  $p < 0.03$ ,  $\eta_p^2 = 0.29$ , in Experiment 1;  $F(1, 31) = 14.58$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.32$ , in Experiment 2; and  $F(1, 31) = 40.02$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.56$ , in Experiment 3. Thus, the advantage in RT that occurred when a target appeared in the rich quadrant rather than a sparse quadrant increased over time. Search accuracy was high in all three experiments (higher than 97%) and was comparable between the rich and sparse conditions ( $p > 0.09$  in all experiments).

The emergence and strengthening of a spatial bias in Experiments 1 to 3 is not likely due to oculomotor learning. In Experiment 4, participants viewed displays presented for approximately 180 ms. By presenting the display briefly, this experiment provided an accuracy rather than RT measure of probability cueing. In addition, the limited display duration significantly reduced the utility of oculomotor learning in task performance because few, if any, saccades could be made in 180 ms. Yet search was still facilitated (this time measured by accuracy) when a target appeared in the rich quadrant, rather than in a sparse quadrant,  $F(1, 21) = 28.40$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.58$ . Moreover, this accuracy advantage increased over time,  $F(1, 21) = 7.64$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.28$ , for the linear trend of the interaction term between quadrant and block.

Having established a spatial bias toward the high-probability, rich quadrant, we next examined whether this effect remained in the same environmental locations or the same visual field locations following the 90° tilt in the participant’s body and/or head.

## Testing phase

In Experiment 1, participants leaned against a slanted surface that tilted their body and head 45° away

from vertical. In the testing phase, they tilted their body and head in the opposite direction, producing a 90° change in viewpoint. This manipulation dissociated the viewer-centered reference frame from the environment-centered reference frame, allowing us to evaluate whether learning is tied to the viewer’s perspective, is referenced relative to the environment, or some combination of the two. Following the change in body orientation, the search target could appear in three types of quadrants (Figure 1A): the screen location that was “rich” during training (the *world-rich* condition), the visual field location that was rich during training (the *viewer-rich* condition), or either of the other two quadrants (the *sparse* condition). Accuracy was high (greater than 96.5%) and equivalent across all conditions in Experiment 1,  $F(2, 30) = 1.18$ ,  $p > 0.30$ . This was also the case in Experiments 2 and 3 ( $F$ ’s  $< 1$ ).

As shown in Figure 3A, search RT in Experiment 1’s testing phase differed significantly among the three testing conditions,  $F(2, 30) = 13.95$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.48$ . Planned contrasts showed that RT was significantly faster in the *world-rich* condition than in the *sparse* condition,  $t(15) = 3.12$ ,  $p < 0.007$ . It was also significantly faster in the *viewer-rich* condition than in the *sparse* condition,  $t(15) = 4.46$ ,  $p < 0.001$ . Finally, the *viewer-rich* condition was significantly faster than the *world-rich* condition,  $t(15) = 2.67$ ,  $p < 0.017$ . The first two comparisons remained significant following a Bonferroni correction for multiple comparisons (critical  $p = 0.0133$ ). These data suggest that viewers use multiple reference frames to learn where to guide attention. The viewer-centered component was significantly stronger than the world-centered component, suggesting that attention may be predominantly viewer centered.

Breaking Experiment 1’s testing phase into eight blocks of trials (Figure 3B) revealed no interaction between condition and block,  $F(14, 210) = 1.24$ ,  $p > 0.25$ , suggesting that the spatial biases developed during the training phase persisted for nearly 200 trials. The long-term persistence of this spatial bias was replicated in subsequent experiments and will not be further reported.

Experiment 1 showed that a durable form of attention remained at the same environmental locations and the same visual field locations as the previously target-rich region. However, it is possible that these findings may be limited to the nature of the tilt manipulation used in Experiment 1. Participants were tilted 45° from vertical in different directions across the training and testing phases of the experiment. Although this resulted in a 90° difference between phases, participants could have aligned attention to the gravitational axis by rotating 45° clockwise or counterclockwise in both phases. Therefore, in Experiment 2, participants were trained while they sat upright, with

their up-down axis aligned with that of the room. They then laid down horizontally in the testing phase (Figure 4). Results replicated those of Experiment 1. As shown in Figure 4, search RT differed significantly across the three testing conditions,  $F(2, 62) = 19.44$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.39$ . Planned contrasts showed that RT was faster in the *world-rich* condition compared with the *sparse* condition,  $t(31) = 3.12$ ,  $p < 0.004$ . It was also faster in the *viewer-rich* condition compared with the *sparse* condition,  $t(31) = 7.01$ ,  $p < 0.001$ . Finally, the *viewer-rich* condition was significantly faster than the *world-rich* condition,  $t(31) = 2.85$ ,  $p < 0.008$ . All of these values exceeded the Bonferonni-corrected alpha threshold.

What served as the basis for the viewer-centered reference frame? Was it the head, the body, or both? In monkeys, neurons in the parietal cortex code visual space using multiple egocentric coordinate systems, including head-centered, body-centered, and eye-centered coordinates (Andersen, Snyder, Bradley, & Xing, 1997). However, previous studies of durable attention did not differentiate between these different egocentric coordinate systems. In Experiment 3, we asked participants to tilt their head  $45^\circ$  in one direction during training and  $45^\circ$  in the other direction during testing. This instruction produced a  $90^\circ$  rotation in the head (and eyes) but did not affect body position (Figure 5).

As shown in Figure 5, search RT differed significantly across the different testing conditions of Experiment 3,  $F(2, 62) = 18.54$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.37$ . All pairwise comparisons were significant according to a Bonferonni-corrected  $p$  value (0.013): faster RT in the *body+world rich* condition than the *sparse* condition, faster RT in the *head-rich* condition than the *sparse* condition, and faster RT in the *head-rich* condition than the *body+world rich* condition, smallest  $t(31) = 3.10$ , largest  $p < 0.004$ . The presence of a strong attentional bias toward the *head-rich* quadrant shows that the viewer-centered representation used in durable attention is strongly head and/or eye centered.

Was there evidence for a body-centered, rather than head-centered, representation? Experiments 1 and 3 differed mainly in whether the participant's body aligned with his or her head (Experiment 1) or not (Experiment 3). If viewer-centered attention is partly body centered, then Experiment 3 should produce a stronger effect in the world-centered rich quadrant than Experiment 1. In addition, the viewer (head)-centered component should be weaker in Experiment 3 than in Experiment 1. Contrary to this prediction, an analysis of variance on testing condition as a within-subject factor and experiment (1 or 3) as a between-subject factor showed no interaction between testing condition and the effect of body and head alignment,  $F < 1$ . Thus, the viewer-centered component was primarily head/eye centered rather than body centered. Future

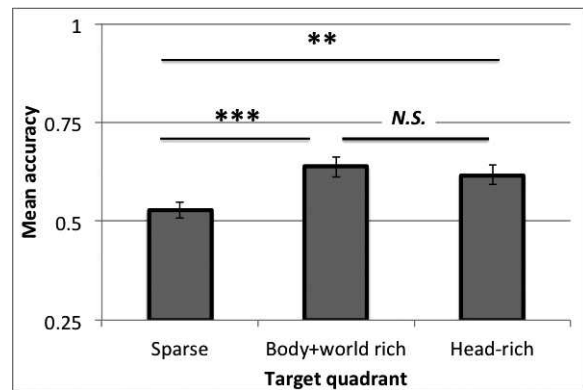


Figure 6. Results from the testing phase of Experiment 4. Error bars show  $\pm 1$  SE of the mean.  $**p < 0.01$ ;  $***p < 0.001$ ; N.S., not significant.

studies that directly examine whether body-centered representations contribute to probability cueing are needed to confirm these results.

The first three experiments revealed the co-existence of an environment-centered and a viewer-centered spatial bias. But did these components depend on oculomotor learning? To find out, in Experiment 4 we replicated Experiment 3 while minimizing saccadic eye movements. The display was presented briefly to limit saccades. Accuracy in the testing phase (Figure 6) significantly varied across the three testing conditions,  $F(2, 42) = 8.41$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.29$ . Planned contrasts showed that search was more accurate in the *body+world rich* condition than in the *sparse* condition,  $t(21) = 4.17$ ,  $p < 0.001$ , and in the *head-rich* condition than in the *sparse* condition,  $t(21) = 3.49$ ,  $p < 0.002$ . The *body+world rich* condition did not differ from the *viewer-rich* condition,  $t(21) = 0.63$ ,  $p > 0.50$ .

In contrast to the previous experiments, the magnitudes of the viewer-aligned and the environment-aligned components of probability cueing were similar in Experiment 4. This suggests that eye movements may have strengthened the viewer-aligned component in the previous experiments. However, the procedure of Experiment 4 differed in several ways from the other experiments. One notable difference is the fact that the items' sizes depended on their screen location, perhaps increasing the salience of the item's location relative to environmental landmarks. Another important consideration is the fact that the distribution of the target may have been learned more poorly when the display was briefly presented. Consistent with this possibility, the number of trials in which participants successfully located the target was reduced (i.e., error rates were high). It is possible that the viewer-aligned and environment-aligned components of location probability learning are acquired at different rates.

Because no eye tracking was used in Experiment 4, we could not rule out the contribution of anticipatory

saccades to performance. However, it is important to note that anticipatory saccades to the rich quadrant would have to be the *outcome* rather than the *cause* of learning the target's locations. Because few, if any, saccades could be made while the display was presented, the opportunity for developing an oculomotor routine during search was limited. As we will see next, participants were generally unaware of where the target-rich region was, lowering the likelihood that they had intended to plan a saccade to the rich region before the display onset.

Finally, we asked whether participants became explicitly aware of where the target was likely to appear. Because each participant made just one mouse click in the recognition test, power considerations required the pooling of data from Experiments 1 to 3 (combined  $n = 80$ ), which yielded qualitatively similar results.

In the testing phase, the target was in the *world-rich* quadrant 25% of the time, the *viewer-rich* quadrant 25% of the time, and the two *sparse* quadrants 50% of the time. If participants had no recoverable knowledge about the target's location, the percentage of participants choosing the three types of quadrant should match these numbers. Instead, 37.5% of the participants chose the *world-rich* quadrant, 27.5% the *viewer-rich* quadrant, and 35% a *sparse* quadrant. These values deviated from chance,  $\chi^2(2) = 8.80$ ,  $p = 0.012$ . Follow-up binomial tests showed that the *world-rich* choice exceeded chance ( $z = 2.58$ ,  $p = 0.005$ ), but the *viewer-rich* choice did not differ from chance ( $z = 0.52$ ,  $p > 0.30$ ). Thus, whereas explicit awareness may have contributed to an attentional bias toward the *world-rich* quadrant, the *viewer-centered* attentional bias was largely implicit.

## Discussion

This study investigated how the rotation of the viewers' bodies and/or heads through the vertical plane influenced where spatial attention was allocated. Four experiments showed that durable attention persists in two regions of space: one that rotates with the viewer and the other that remains in the same environmental locations as the previously attended locations.

These findings are reminiscent of those from patients with hemifield neglect. When asked to lie down on one side of their body, neglect patients neglect the left side of space relative to their body and the left side of the environment assuming an upright posture (Calvanio, Petrone, & Levine, 1987; Farah, Brunn, Wong, Wallace, & Carpenter, 1990). Farah et al. (1990) propose that spatial attention is coded relative to both the viewer-centered reference frame and the environment-centered reference frame. Our data support this

conclusion but further indicate that these different reference frames are used by neurologically normal individuals, can develop rapidly, and can be acquired under conditions of incidental statistical learning. In addition, because Farah et al. (1990) tested only whole-body rotation, they could not determine whether the viewer-centered reference frame is head centered or body centered. In contrast, we tested both whole-body rotation and head-only rotation. Our results showed that the viewer-centered representation depends primarily on head orientation rather than on body orientation. This finding opens new opportunities for future neurophysiological, eye tracking, and computational modeling studies that test the viewer-centered representation that underlies persistent forms of attention. Of course, this finding does not imply that body position is irrelevant to all forms of spatial attention. In other studies, body position, such as the position of the participants' hands, modulates attention (Abrams, Davoli, Du, Knapp, & Paull, 2008; Davoli & Brockmole, 2012).

Our results both relate to and differ from previous investigations on the coordinate systems of transient forms of attention. Owing to the rapid nature of spatial attention effects, previous studies have been limited to introducing eye movements between an attentional cue and the subsequent measurement of its effects on perception (Abrams & Pratt, 2000; Golomb et al., 2008; Harrison et al., 2012; Mathôt & Theeuwes, 2010; Maylor & Hockey, 1985; Pertzov et al., 2010; Pertzov et al., 2011; Posner & Cohen, 1984). Unlike our study, most others have reported that attentional enhancements are coded either retinotopically or spatiotopically. When evidence for both reference frames is observed in a study, the experimental conditions that give rise to them are often reported to differ (e.g., they have different time courses or are evident with different motor effectors; Abrams & Pratt, 2000; Mathôt & Theeuwes, 2010). Like Farah et al. (1990), our study represents one of the few cases in which strong attentional effects are found in both the viewer-centered and the environment-centered reference frames and under the same experimental conditions. In addition, whereas previous findings of a spatiotopically coded attentional bias could be alternatively accounted for by a body- or head-centered representation, our study cleanly dissociates the viewer from the environment. Nonetheless, the viewer-centered component may be similar to retinotopically based attentional biases that have been reported by others (e.g., Golomb et al., 2008; Golomb & Kanwisher, 2012). In particular, it may be related to the mechanism that supports the form of persistent attention studied here. We have previously proposed that this form of attention, probability cueing, reflects learning the direction that attention should be shifted to find a target (Jiang,



Swallow, & Capistrano, 2013). The close correspondence between spatial attention shifts and eye movements suggests that an eye-centered reference frame may be employed in persistent attention. Future studies that manipulate eye position are needed to dissociate retinotopic from head-centered representations in persistent attentional biases.

The observation of an environment-centered component, although intuitive, differs substantially from our previous studies that introduced viewpoint changes through viewer locomotion. In those studies, participants performed the visual search task on a display that laid flat on the table. After the training phase, participants stood up, moved 90° to a new seating position at another side of the display, and completed the testing phase. Participants found the target faster in the viewer-rich quadrant than in the sparse quadrants. In contrast to the current study, however, no search advantage was found in the world-rich quadrant (Jiang & Swallow, 2013). In fact, an environment-centered spatial bias failed to develop even with the addition of an unvarying natural scene in the background of visual search (Jiang, Swallow, & Sun, in press).

Thus, a critical factor in determining whether spatial attention remains in an environment-centered reference frame appears to be the manner in which the viewpoint change is introduced. Spatial attention lingers at the world-rich region when the participants tilt their body/head through the vertical plane but not when they walk from one edge of the table to the other. In both cases, plenty of environmental cues are available for establishing viewpoint invariance and for spatial updating, including the room layout, furniture and other landmarks, and the monitor itself. These cues therefore do not appear to be sufficient for establishing an environment-centered representation. Rather, we propose that the environment-centered representation survives body/head tilt because the vertical (gravitational) axis is used to encode or update the visual search display.

This proposal is consistent with research on object recognition and mental rotation. In a series of studies, Rock found that the gravitational axis is critical for shape identification (reviewed in Rock, 1997). After learning novel shapes while sitting upright, participants fail to recognize them when the shapes are rotated on the screen by 90° in the vertical plane. This is true even when participants tilt their head to bring the shape to its original orientation on their retina. Rock proposes that the “top” of the shape is aligned with the gravitational axis rather than with the retina. Similarly, mental rotation is most successful when the rotational axis aligns with the gravitational axis (Asakura & Inui, 2011; Corballis, Zbrodoff, & Roldan, 1976; Waszak, Drewing, & Mausfeld, 2005). Other data indicate that eye movements are referenced relative to the horizon, which

could reflect a tendency to use the gravitational axis in representations of the external environment (Cristino & Baddeley, 2009). In our study, coding the display relative to the gravitational axis allows the participant to extract invariant features of a visual display independent of his or her body/head orientation. In contrast, when an upright viewer walks around the gravitational axis (as when one walks around a table), the ability to represent the rich regions relative to the environment depends on spatial updating or a landmark-centered coding of space. These appear to be insufficient to yield an environment-centered attentional bias (Jiang & Swallow, 2013; Jiang, Swallow, & Sun, in press).

The effect of introducing different types of viewpoint changes on the presence of environmentally stable representations also has precedence in hemifield neglect. As reviewed earlier, neglect patients appear to use both viewer-centered and environment-centered reference frames (Farah et al., 1990). However, neglect also appears to be viewer centered when an upright patient changes perspective. Bisiach and Luzzatti (1978) asked two neglect patients to imagine being in a well-known square in Milan. When asked to imagine the square from the perspective facing the cathedral, the patients described what was on the right side of the square from that perspective. But when asked to imagine standing with their back against the cathedral, the patients described the other side of the square, formerly neglected but now on the right side of the mental image. Thus, neglect appears to be viewer centered when the patients imagine a perspective change but contains both viewer- and environment-centered representations when the patients tilt their body through the vertical plane. These data are consistent with our findings.

Several mechanisms have been proposed to account for the ability to maintain visual stability across saccades and therefore might also play a role in the environmentally centered attentional biases reported here (Cavanagh et al., 2010; Wurtz, 2008). One mechanism is “receptive field remapping,” in which a neuron changes its receptive field in anticipation of an impending saccade (Duhamel, Colby, & Goldberg, 1992). Any remapping in our study is likely driven by a shift in attention, rather than, or in addition to, a shift in receptive fields (Cavanagh et al., 2010). However, remapping is a transient mechanism. Yet in our study, several minutes of interruption occurred between the training and testing phases. It is unlikely that receptive field remapping could have persisted under these conditions. Therefore, it seems doubtful that remapping supports the environment-centered representation in our study (Burr & Morrone, 2012). A second possibility is that the attended locations are represented relative to the external environment, including landmarks in the room or other forms of spatiotopic

representation (Burr & Morrone, 2012). We believe that the landmark-based coding is a plausible account of our data, although its success may be limited to viewpoint changes produced by rotations through the vertical plane.

## Conclusion

In this study, we examined the coordinate systems used to code durable forms of spatial attention. Participants first acquired a spatial bias toward one region of the display. Following a rotation of the participants' body and/or head through the vertical plane, spatial attention lingers at locations defined both by the environment-centered reference frame and the viewer-centered reference frame. The viewer-centered component is primarily head (and/or eye) centered, with little contribution from the body-centered coordinates. It may be supported by retinotopic or head-centered computations of space. The environment-centered component may result from coding the visual display based on environmental landmarks, although the success of environment-based coding may be limited to situations involving body and/or head tilt. Future research should examine whether the viewer-centered component is retinotopic or head centered and how explicit awareness affects the coordinate systems of spatial attention.

*Keywords:* attention, spatial reference frame, visual search, visual statistical learning

## Acknowledgments

Daniel Cao, Chris Capistrano, Julia Cistera, Andrew Mekhail, Tayla Smith, Liwei Sun, and Josh Tisdell tested participants. Chris Capistrano and Bo-Yeong Won contributed to artwork. Josh Tisdell helped construct experimental equipment. This work was supported by funds from the University of Minnesota.

Commercial relationships: none.

Corresponding author: Yuhong V. Jiang.

Email: [jiang166@umn.edu](mailto:jiang166@umn.edu).

Address: Department of Psychology, University of Minnesota, Minneapolis, MN, USA.

## References

- Abrams, R. A., Davoli, C. C., Du, F., Knapp, W. H., III, & Paull, D. (2008). Altered vision near the hands. *Cognition*, *107*, 1035–1047. doi:10.1016/j.cognition.2007.09.006.
- Abrams, R. A., & Pratt, J. (2000). Oculocentric coding of inhibited eye movements to recently attended locations. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 776–788.
- Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, *20*, 303–330. doi:10.1146/annurev.neuro.20.1.303.
- Asakura, N., & Inui, T. (2011). Disambiguation of mental rotation by spatial frames of reference. *i-Perception*, *2*, 477–485. doi:10.1068/i0425.
- Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, *14*, 129–133.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Burr, D. C., & Morrone, M. C. (2012). Constructing stable spatial maps of the world. *Perception*, *41*, 1355–1372.
- Calvanio, R., Petrone, P. N., & Levine, D. N. (1987). Left visual spatial neglect is both environment-centered and body-centered. *Neurology*, *37*, 1179–1183.
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, *57*, 1241–1261.
- Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*, 147–153. doi:10.1016/j.tics.2010.01.007.
- Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 224–234.
- Corballis, M. C., Zbrodoff, J., & Roldan, C. E. (1976). What's up in mental rotation? *Perception & Psychophysics*, *19*, 525–530. doi:10.3758/BF03211221.
- Cristino, F., & Baddeley, R. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition*, *17*, 880–903. doi:10.1080/13506280902834696.
- Davoli, C. C., & Brockmole, J. R. (2012). The hands shield attention from visual interference. *Attention, Perception & Psychophysics*, *74*, 1386–1390. doi:10.3758/s13414-012-0351-7.
- Druker, M., & Anderson, B. (2010). Spatial probability AIDS visual stimulus discrimination. *Frontiers in*

- Human Neuroscience*, 4. doi:10.3389/fnhum.2010.00063.
- Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255(5040), 90–92.
- Farah, M. J., Brunn, J. L., Wong, A. B., Wallace, M. A., & Carpenter, P. A. (1990). Frames of reference for allocating attention to space: Evidence from the neglect syndrome. *Neuropsychologia*, 28, 335–347.
- Geng, J. J., & Behrmann, M. (2002). Probability cuing of target location facilitates visual search implicitly in normal participants and patients with hemispatial neglect. *Psychological Science*, 13, 520–525.
- Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *Journal of Neuroscience*, 28, 10654–10662. doi:10.1523/JNEUROSCI.2525-08.2008.
- Golomb, J. D., & Kanwisher, N. (2012). Retinotopic memory is more precise than spatiotopic memory. *Proceedings of the National Academy of Sciences, USA*, 109, 1796–1801. doi:10.1073/pnas.1113168109.
- Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2012). Pre-saccadic shifts of visual attention. *PloS One*, 7, e45670. doi:10.1371/journal.pone.0045670.
- Jiang, Y. V., & Swallow, K. M. (2013). Spatial reference frame of incidentally learned attention. *Cognition*, 126, 378–390. doi:10.1016/j.cognition.2012.10.011.
- Jiang, Y. V., Swallow, K. M., & Capistrano, C. G. (2013). Visual search and location probability learning from variable perspectives. *Journal of Vision*, 13(6):13, 1–13, <http://www.journalofvision.org/content/13/6/13>, doi:10.1167/13.6.13. [PubMed] [Article]
- Jiang, Y. V., Swallow, K. M., Rosenbaum, G. M., & Herzog, C. (2013). Rapid acquisition but slow extinction of an attentional bias in space. *Journal of Experimental Psychology. Human Perception and Performance*, 39, 87–99. doi:10.1037/a0027611
- Jiang, Y. V., Swallow, K. M., & Sun, L. (in press). Egocentric coding of space for incidentally learned attention: Effects of scene context and task instructions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. doi:10.1037/a0033870.
- Mathôt, S., & Theeuwes, J. (2010). Gradual remapping results in early retinotopic and late spatiotopic inhibition of return. *Psychological Science*, 21, 1793–1798. doi:10.1177/0956797610388813.
- Maylor, E. A., & Hockey, R. (1985). Inhibitory component of externally controlled covert orienting in visual space. *Journal of Experimental Psychology. Human Perception and Performance*, 11, 777–787.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Pertsov, Y., Avidan, G., & Zohary, E. (2011). Multiple reference frames for saccadic planning in the human parietal cortex. *Journal of Neuroscience*, 31, 1059–1068. doi:10.1523/JNEUROSCI.3721-10.2011.
- Pertsov, Y., Zohary, E., & Avidan, G. (2010). Rapid formation of spatiotopic representations as revealed by inhibition of return. *Journal of Neuroscience*, 30, 8882–8887. doi:10.1523/JNEUROSCI.3986-09.2010.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 551–556). Hillsdale, NJ: Erlbaum.
- Rock, I. (1997). *Indirect perception*. Cambridge, MA: MIT Press.
- Umemoto, A., Scolari, M., Vogel, E. K., & Awh, E. (2010). Statistical learning induces discrete shifts in the allocation of working memory resources. *Journal of Experimental Psychology. Human Perception and Performance*, 36, 1419–1429. doi:10.1037/a0019324.
- Waszak, F., Drewing, K., & Mausfeld, R. (2005). Viewer-external frames of reference in the mental transformation of 3-D objects. *Perception & Psychophysics*, 67, 1269–1279.
- Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research*, 48, 2070–2089. doi:10.1016/j.visres.2008.03.021.