

Visual search and location probability learning from variable perspectives

Yuhong V. Jiang

Department of Psychology & Center for Cognitive Sciences, University of Minnesota, Minneapolis, MN, USA



Khena M. Swallow

Department of Psychology & Center for Cognitive Sciences, University of Minnesota, Minneapolis, MN, USA



Christian G. Capistrano

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



Do moving observers code attended locations relative to the external world or relative to themselves? To address this question we asked participants to conduct visual search on a tabletop. The search target was more likely to occur in some locations than others. Participants walked to different sides of the table from trial to trial, changing their perspective. The high-probability locations were stable on the tabletop but variable relative to the viewer. When participants were informed of the high-probability locations, search was faster when the target was in those locations, demonstrating *probability cuing*. However, in the absence of explicit instructions and awareness, participants failed to acquire an attentional bias toward the high-probability locations even when the search items were displayed over an invariant natural scene. Additional experiments showed that locomotion did not interfere with incidental learning, but the lack of a consistent perspective prevented participants from acquiring probability cuing incidentally. We conclude that spatial biases toward target-rich locations are directed by two mechanisms: incidental learning and goal-driven attention. Incidental learning codes attended locations in a viewer-centered reference frame and is not updated with viewer movement. Goal-driven attention can be deployed to prioritize an environment-rich region.

experimental paradigm for studying the nature of spatial attention (Treisman, 1988; Wolfe, 1994, 2007). Most studies in the lab present the search target at random locations on the display, but in the real world, the locations of search targets are often constrained by their context (Biederman, 1972; Biederman, Mezzanotte, & Rabinowitz, 1982). For example, a mailbox is often found on the side of the street rather than in the driveway. Laboratory studies have examined the impact of such statistical regularities on human performance. The general findings are that humans are highly sensitive to statistical regularities. For example, in contextual cuing, participants are faster to find a target within configurations that occasionally repeat (Brady & Chun, 2007; Chun & Jiang, 1998). In probability cuing, participants are faster to find a target in locations that frequently contained the target before (Geng & Behrmann, 2002, 2005; Jiang, Swallow, Rosenbaum, & Herzig, 2013; Miller, 1988). Statistical regularities modulate spatial attention. Contextual cuing increases the magnitude of *N2pc*, a component of the event-related potential that indexes spatial attention (Johnson, Woodman, Braun, & Luck, 2007). Probability cuing results in reduced search slope—less time is needed per item when the target is in the high-probability locations (Jiang, Swallow, & Rosenbaum, 2013). Knowledge, including that acquired from implicit learning, serves as a powerful cue for spatial attention (Chun, 2000).

However, visual search studies have rarely examined the coordinate system used to code target locations. The vast majority of studies in the lab test stationary participants in front of a static visual display. When a

Introduction

Many daily tasks involve visual search, such as looking for a friend at the airport or snatching a drink from the fridge. Visual search is also a useful

Citation: 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

participant frequently finds the target in certain screen locations, it is unclear whether the attended locations are coded relative to the external world or relative to the viewer. This question is important because the reference frame of attention affects how and where attention is allocated in the future. Coding attended locations in an allocentric (environment-centered) reference frame may lead one to favor those locations after a perspective change, but coding attended locations in an egocentric (viewer-centered) reference frame may lead one to favor new locations after movements through space.

Several theories about spatial attention and statistical learning can be used to generate predictions about the spatial reference frame of implicitly learned attention. The first possibility is that attended locations are coded allocentrically, relative to the external world. Consider the broader problem of vision. All visual input must first be coded by photoreceptors on the retina, but perception often corresponds to the distal stimuli (i.e., what is out there) rather than the proximal stimuli (i.e., what is on the retina; Palmer, 1999; Rock, 1985). The pressure to discover the structure of the external world may apply with equal force to attention, especially the kind of attention acquired through repeated experience with an environment. Implicitly learned attention occurs over repeated interactions with an environment. The acquired knowledge reflects the statistical regularities of the environment. Consequently, attended locations may be coded relative to environmental cues.

Two other possibilities exist, both of which involve the viewer-centered coding of attended locations. According to these accounts, frequently attended locations are coded relative to the viewer's head and/or body. These accounts may seem implausible due to their limitations. In an environment where the viewers move around, coding attended locations relative to the viewer is not optimal for extracting visual regularities. However, a viewer-centered representation can mimic an environment-centered representation provided that visual space is updated during viewer movement (Wang, 2012; Wang & Simons, 1999). Suppose someone places a cup on the table and codes its location egocentrically, as being on her left. Suppose she then walks to the opposite side of the table so the cup is now on her right. Even though she had previously coded the cup's location relative to herself, she would not look to her left to find it. This is because as she walks, visual and proprioceptive cues tell her that she has moved, and hence the spatial relationship of the cup relative to her has changed. As long as she successfully updates the spatial relationship, a viewer-centered representation can lead to stable representations of the world (Simons & Wang, 1998; Wang & Simons, 1999). Thus, viewer-centered coding plus

successful spatial updating can yield a pattern of performance that mimics an environment-centered representation. However, the lack of spatial updating combined with a viewer-centered representation would yield an attentional bias that moves with the viewer.

To summarize, three possibilities exist regarding the spatial representation of incidentally learned attention: people could represent attended locations relative to the external world (*environment centered*); they could represent attended locations relative to their head or body and perform spatial updating as they move (*viewer centered with spatial updating*); or they could code attended locations egocentrically without spatial updating (*viewer centered without spatial updating*).

Recent studies have examined the spatial reference frame of attention but most focus on transient forms of attention that change from trial to trial. These studies demonstrate that attention uses multiple frames of reference. Many phenomena exhibit an egocentric component, including inhibition of return (Abrams & Pratt, 2000; Mathot & Theeuwes, 2010), negative priming (Tipper, Howard, & Houghton, 1998), spatial memory (Golomb, Chun, & Mazer, 2008; Golomb, Pulido, Albrecht, Chun, & Mazer, 2010), and priming of pop-out (Ball, Smith, Ellison, & Schenk, 2009, 2010). Other studies have uncovered allocentric representations, including spatiotopic coding of locations or object-centered attention (Ball et al., 2009; Behrmann & Tipper, 1999; Golomb et al., 2010; Mathot & Theeuwes, 2010; Pertzov, Zohary, & Avidan, 2010; Posner & Cohen, 1984). Studies in psychophysics and neurophysiology have demonstrated the phenomenon of receptive field remapping: a neuron is activated by a stimulus outside its receptive field if an impending saccade will bring it into the receptive field (Cavanagh, Hunt, Afraz, & Rolfs, 2010; Nakamura & Colby, 2002; Wurtz, 2008). Such remapping could support visual stability across eye movements. However, because these studies involve transient forms of attention, they do not generate clear predictions about what happens with more durable forms of attention. They are also unable to dissociate environment from body and head-centered representation.

Two major differences exist between transient forms of attention and incidentally learned attention. First, transient forms of attention are typically not acquired through statistical learning; they do not reflect the structure of the external world, but rather reflect goals and salience. Second, because implicitly learned attention lasts long after the initial training is complete (Chun & Jiang, 2003; Jiang, Swallow, Rosenbaum, & Herzig, 2013), there is sufficient time for viewers to move. Viewer movement is important for successful spatial updating (Rieser, 1989; Simons & Wang, 1998; Tsuchiai, Matsumiya, Kuriki, & Shioiri, 2012). If attention is viewer centered, the likelihood that its representation

is successfully updated increases when the viewer moves rather than when the display rotates (Simons & Wang, 1998; Tsuchiai et al., 2012). Consequently, one may predict that visual statistical learning will yield either an environment-centered representation, or a successfully updated viewer-centered representation.

However, a recent study provides no evidence for the above prediction (Jiang & Swallow, 2013). In that study, participants sat at one side of the table and performed visual search on a monitor laid flat on the table. They searched for a rotated T among rotated Ls and reported the T's color. Unbeknownst to them, across multiple trials the T was more often found in one, rich quadrant (50% probability) than in any one of the sparse quadrants (16.7% probability). After 384 trials, participants moved their chair to another side of the table, producing a 90° change in perspective. In addition, the target was now randomly placed, such that it appeared in each quadrant 25% of the time. Results showed that participants acquired probability cuing in the training phase. The spatial bias persisted for nearly 200 trials of testing. If probability cuing is environment centered, then the attentional bias should have persisted in the original screen locations. But this was not the case: in the testing phase, the spatial bias was directed toward the region of the display that maintained the same spatial relationship with the viewer as the previously rich quadrant. Additional experiments replicated these results with briefly presented displays (Jiang & Swallow, 2013) and when visual search was conducted against a natural scene (Jiang, Swallow, & Sun, 2013). These data do not support the idea that attended locations are coded relative to the environment, or that attended locations are viewer centered but are updated with viewer movement.

The findings of Jiang and Swallow (2013) are surprising and perplexing. They imply that there exist substantial limits to human attention and learning systems. Because attended locations are coded relative to the viewer and not updated after viewer movement, this learning mechanism seems to be of limited utility. For example, it would not be a powerful mechanism for foraging or spatial navigation. With an egocentric representation and no spatial updating, one may fail to find food-rich locations in the forest if search starts from random locations. An egocentric system without spatial updating is useful only in situations in which navigation or foraging paths are highly constrained.

The present study examines visual search and location probability learning in observers who assume variable perspectives. All of the experiments involved participants who moved from trial to trial around a tabletop. To mimic foraging, the target was frequently located in a specific quadrant of the display. Due to viewer movement, this quadrant was random relative to

the viewer but fixed in the environment. This design led to an important difference from the Jiang and Swallow (2013) study, which trained and tested stationary viewers. Because the target-rich quadrant was random relative to the viewer, no consistent associations could be built between the target-rich quadrant and the viewer. In the absence of strong egocentric coding of attended locations, this experiment may have unmasked or encouraged the learning of the environment-centered attentional bias. With this design, we tested location probability cuing under incidental learning and intentional learning conditions. In previous studies, we have argued that incidentally learned attention differs fundamentally from goal-driven attention (Jiang, Swallow, & Rosenbaum, 2013). If this is true, then top-down knowledge about the target's location probability may change the nature of spatial coding.

This study addressed several novel questions. First, it characterized visual search behavior in observers who moved from trial to trial. This departed from the vast majority of visual search studies that tested stationary observers. Second, it contrasted the intentional allocation of attention with incidental learning, dissociating two sources of top-down knowledge. Third, the study provided constraints on the ubiquity of visual statistical learning. As is shown, consistent visual statistics were necessary, but not sufficient, for learning. The study findings have implications for theories of visual search and spatial attention.

Methods

Participants

Sixty-four college students (17 men and 47 women, 18–35 years old) participated in this study. There were 16 participants in each of the four experiments. Two additional participants were tested, but their data were removed due to a failure to follow instructions (see Design). All participants had normal or corrected to normal visual acuity and were naïve to the purpose of the study. 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 the experiment. Participants received \$10/hour or extra credit for their participation.

Apparatus

Participants conducted visual search in a room with dim lighting. A 17-inch touch screen monitor (75 Hz vertical refresh rate; 1024 × 768 pixels) was laid flat on

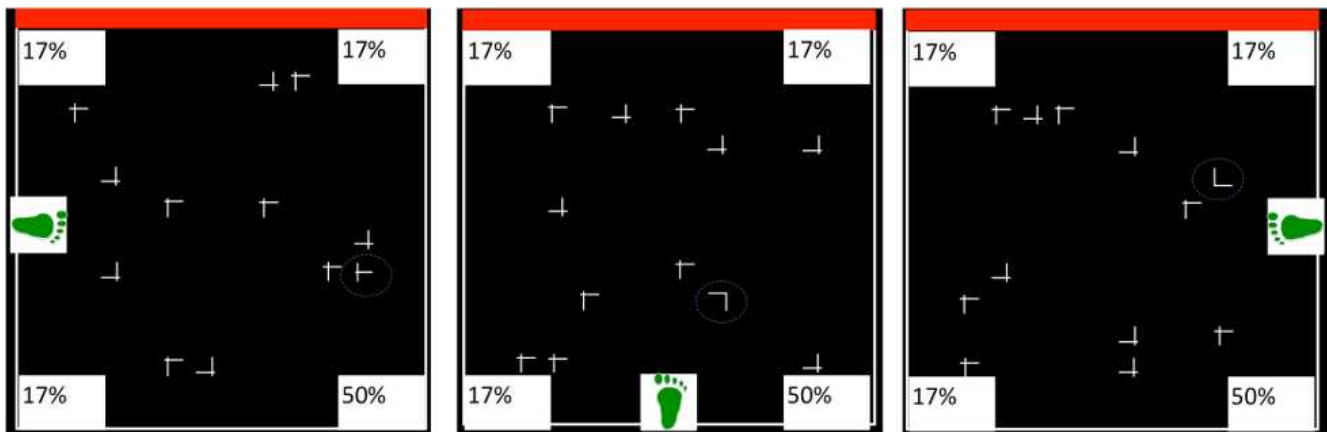


Figure 1. Sample search displays on three trials used in Experiment 1. The thick red bar was constantly displayed to provide landmark information. Green footprints indicated where participants should stand (see Methods). Participants clicked the left mouse button for a T and the right mouse button for an L. Some information is shown here for illustrative purposes only and was not actually displayed (i.e., the dotted circle around the target, percentages regarding the target's location probability, and trial number). These three trials illustrate a rich quadrant that is stable on the monitor but variable relative to the viewer.

a 38-inch-tall stand. Tape on the floor marked four equidistant positions (*bases*) around the monitor, to which participants moved as indicated by a green footprint icon on the monitor (Figure 1). Participants used a wireless mouse to make responses. They held the wireless mouse in their dominant hand throughout the experiment, including the locomotion period. In addition to the monitor and stand, other furniture in the room acted as environmental landmarks. The experiments were programmed with Psychtoolbox (Brainard, 1997; Pelli, 1997) implemented in MATLAB (www.mathworks.com).

Stimuli

Experiment 1

Each display (Figure 1) contained 12 items placed in randomly selected locations in an invisible 10 by 10 matrix (19 by 19 cm). Viewing distance varied according to the participant's height but was estimated to be between 55 and 85 cm. Due to the variability in participants' height (and hence viewing distance), we report stimulus size in centimeters rather than in visual angles. At a viewing distance of 57 cm, 1 cm on the display is approximately 1° visual angle. There were three items in each visual quadrant. The target was either a T or an L (1.3 by 1.3 cm). The distractors were distorted plus symbols (+). The high similarity between the targets and distractors led to relatively long response times (RT), which gave participants plenty of time to learn the visual statistics used in this study. To ensure that the participants' standing position did not affect how the items appeared, all items had a random orientation of 0° , 90° , 180° , or 270° . The entire search

region was framed by a white square. A fixed side of the square was marked with a thick red bar (19 by 1.3 cm) to provide a constant landmark. Participants made a left mouse click for a T or a right mouse click for an L. Both accuracy and speed were emphasized.

Experiment 2

This experiment used the same stimuli as Experiment 1, except that a scene was displayed as the background of visual search. The scene was randomly selected for each participant but remained the same throughout the experiment. To increase the visibility of the search items, all items were placed inside a small gray circle (1.8 by 1.8 cm). Figure 2 shows a sample search display on two consecutive trials of Experiment 2.

Experiment 3

The same stimuli from Experiment 1 were used in Experiment 3. However, one side of the monitor was designated as the "home base," to which participants always returned for visual search (see Design).

Experiment 4

This experiment used the same stimuli as Experiment 1, except that participants received explicit instructions about where the target was likely to be (see Design).

Design

Using a probability cuing paradigm, we manipulated the target's location probability. The target appeared in



Figure 2. Sample displays used in three consecutive trials of Experiment 2. The experiment was identical to Experiment 1 except that a scene was placed in the background. The green footprint indicated participants' standing positions. The scene remained the same throughout the experiment. The rich quadrant was at a fixed part of the scene independent of where the viewer stood.

a “rich” quadrant on 50% of the trials and in any one of three “sparse” quadrants on 16.7% of the trials. Exactly which quadrant was rich was counterbalanced across participants but remained the same for a given participant. Learning was incidental in Experiments 1–3 but intentional in Experiment 4.

In all experiments, the rich quadrant was fixed on the monitor. Participants stood at a randomly determined base on each trial of Experiment 1 (“incidental learning experiment”), so the rich quadrant was variable relative to their viewpoint. For example, the rich quadrant might be to their lower left on one trial, upper right on another, and so on (Figure 1). This setup resembles many real-world foraging situations, in which the high-reward locations are stable in the environment but can be approached from different directions.

Experiment 2 (“incidental learning with scene”) was the same as Experiment 1, except that search was conducted against a natural scene that remained the same for all trials (Figure 2). Therefore, the rich quadrant was also stable relative to the scene.

In Experiment 3 (“incidental learning with home base”) participants always performed visual search from a single home base. However, before each trial, they walked halfway toward a randomly selected position (any one of four bases around the monitor), and walked back to the home base. This experiment produced the same amount of viewer movement as the other experiments. However, the rich quadrant was fixed on the monitor, so it was stable relative to both the viewer and the environment during search.

We did not inform participants of the target's location probability in Experiments 1–3. In Experiment 4 (“intentional learning”) participants received explicit instructions about where the target was likely to appear. This experiment was identical to Experiment 1, except that at the beginning of the experiment the

computer program informed participants of the target-rich quadrant. A blue square the size of a visual quadrant was placed on the target-rich quadrant. Participants were told that the target would appear in that region in 50% of the trials, and in each of the other three quadrants in 17% of the trials. They were encouraged to prioritize search in the rich quadrant. In addition, they were told that the target-rich quadrant remained in the same place on the monitor, regardless of where they stood. An experimenter verbally reinforced these instructions and gave participants a reminder every 96 trials. Despite this instruction two participants reported that they ignored it, so their data were replaced.

Procedure

Each trial started with a green footprint placed at a random side of the monitor (base). Participants moved to the base indicated by the footprint in Experiments 1, 2, and 4, or moved halfway toward it and returned to the home base in Experiment 3. An experimenter stayed in the room to ensure that participants followed these instructions. Once in position, participants touched a central fixation point to initiate the trial, which erased the footprint. The touch response required eye–hand coordination and ensured that eye position moved away from the previous target location. After a 300-ms interval, an array of items appeared and remained on the display until participants responded with the wireless mouse. The target (T or L) was randomly determined on each trial, so the corresponding motor response was also randomized.

The main phase of the experiment had 384 trials (16 blocks of 24 trials each), during which participants moved from trial to trial. All experiments ended with a

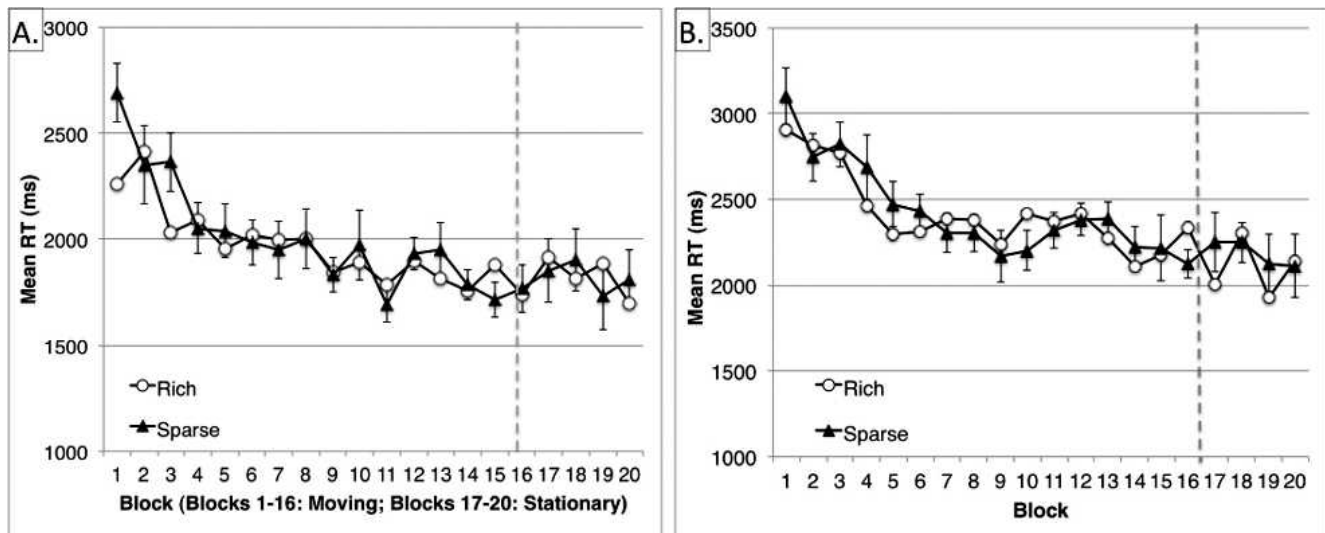


Figure 3. Results from Experiments 1 (A) and 2 (B). Mean RT as a function of target quadrant type (rich or sparse) and experimental block. Each block had 24 trials. Blocks 1–16: The target was more likely to appear in the rich quadrant and participants' search position was variable. Blocks 17–20: The target was equally likely to appear in all quadrants and participants did not move. Error bars show ± 1 SE of the difference between rich and sparse conditions.

stationary phase (four blocks of 24 trials) to test whether cuing was evident in the absence of locomotion. In this phase participants always stood at the home base and did not move between trials. To prevent new learning the target was equally likely to appear anywhere on the screen during the stationary phase. Because incidentally acquired probability cuing persists beyond initial training (Jiang, Swallow, Rosenbaum, & Herzig, 2013), any latent learning acquired while participants were moving should persist in the stationary phase. In Experiment 4, before the stationary phase participants were told that the target would now be randomly placed and that they should abandon any systematic prioritization of any part of the display.

To assess explicit awareness of learning, at the end of the experiment, participants stood at the home base and touched the quadrant where they believed the target was most often found.

Results

Experiment 1. Incidental learning

Accuracy in the rich (98.0%, $SE = 0.4\%$) and sparse quadrants (98.0%, $SE = 0.4\%$) was statistically comparable, $t < 1$. Figure 3A shows mean RT, excluding incorrect trials and trials with RTs longer than 9.5 s. The percentage of trials removed due to very long RT

(>9.5 s) was 0.22%, 0.71%, 0.28%, and 0.23% for Experiments 1 to 4, respectively.

In the locomotion phase (Blocks 1–16) RT showed a significant improvement as the experiment progressed, $F(15, 225) = 9.53$, $p < 0.001$, $\eta_p^2 = 0.39$ for the main effect of block. However, neither the main effect of target quadrant, $F(1, 15) = 1.93$, $p > 0.18$, nor the interaction between target quadrant and block, $F(15, 225) = 1.53$, $p > 0.11$, was significant.

When observers became stationary (Blocks 17–20), probability cuing remained absent, $F < 1$ for the main effect of quadrant condition, and $F < 1$ for the interaction between quadrant condition and block. Thus, a stable environment alone was insufficient for developing probability-guided attentional search. Unequal location probability was not learned when the rich target quadrant was variable relative to the viewer.

Experiment 2. Incidental learning with scene

Experiment 2 replicated Experiment 1 (Figure 3B). For this experiment, items were displayed over a constant natural scene, so the target-rich locations could be coded relative to the background scene. Nonetheless, probability cuing did not emerge. Accuracy was statistically comparable between the rich (97.2%, $SE = 0.8\%$) and the sparse quadrants (96.8%, $SE = 0.7\%$), $t < 1$. In the mobile phase (Blocks 1–16), search RT became faster as the experiment progressed, $F(15, 225) = 8.79$, $p < 0.001$, $\eta_p^2 = 0.37$. However, neither the main effect of target quadrant nor the target

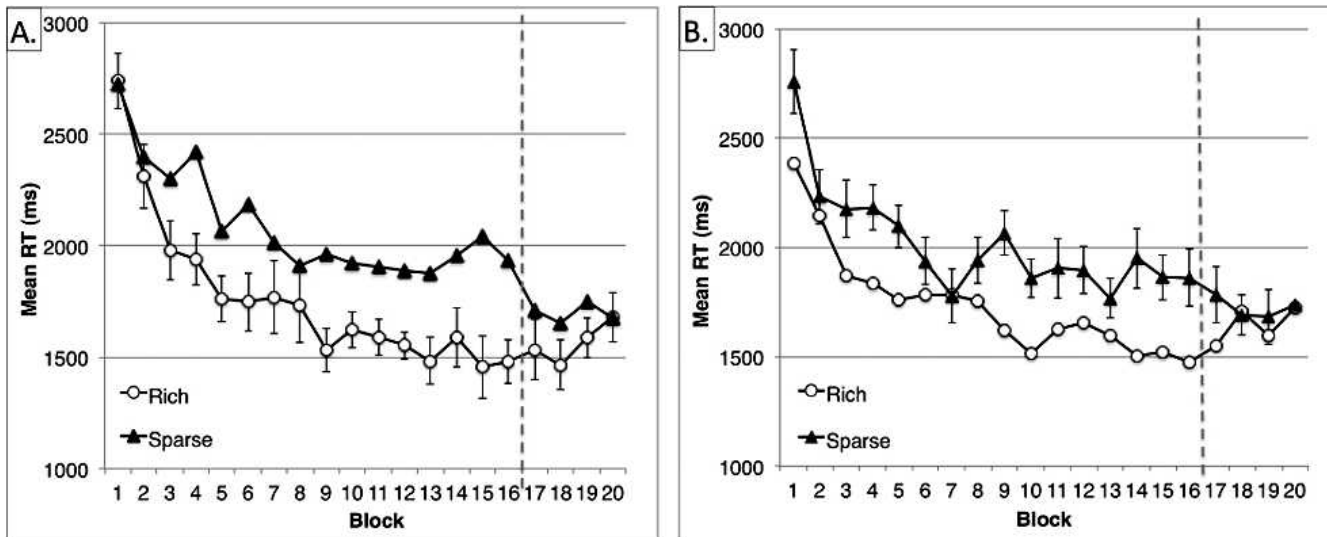


Figure 4. Results from Experiments 3 (A) and 4 (B). See Figure 3 for information about experimental blocks and error bars.

quadrant by block interaction was significant, $F_s < 1$. The stationary phase (Blocks 17–20) also revealed no main effect or interaction of target quadrant, $F_s < 1$. Thus, when attended locations were unstable relative to the participants, probability cuing was not acquired even when the target was often found in one quadrant of the display and when the search was conducted against an invariant natural scene.

Experiment 3. Incidental learning with home base

The lack of probability cuing in the first two experiments could have arisen from two possibilities. First, attended locations may have been coded in an egocentric reference frame that was not updated by viewer movement. Because the high-probability regions were random relative to the viewer, learning was disrupted. Alternatively, it is possible that viewer locomotion between trials interfered with learning. To rule out the second possibility, in Experiment 3 participants moved the same amount as in Experiments 1 and 2 but always searched from the home base. They walked halfway toward another base and walked back to the home base. Although participants moved between trials, search was always conducted from the same home base. Thus, the target-rich regions were not only stable in the external environment, but also stable relative to the viewer. If locomotion itself interferes with learning, then probability cuing should have been absent in this experiment.

Results in Experiment 3 showed significant learning (Figure 4A). Participants were more accurate when the target was in the rich quadrant (98.1%, $SE = 0.5\%$) than in the sparse quadrants (97.0%, $SE = 0.6\%$), $t(15) =$

3.32, $p < 0.005$, Cohen's $d = 1.71$. In addition, in the mobile phase (Blocks 1–16), RT was significantly faster in the rich quadrant than in sparse quadrants, $F(1, 15) = 33.51$, $p < 0.001$, $\eta_p^2 = 0.69$ for the main effect of target quadrant, and this difference increased as the experiment progressed, $F(15, 225) = 1.88$, $p < 0.026$, $\eta_p^2 = 0.11$, for the interaction between target quadrant and block. When observers became stationary and the target was randomly placed (Blocks 17–20), probability cuing persisted in the first three stationary blocks, $F(1, 15) = 3.49$, $p < 0.04$ (one-tailed), then dissipated in the last block, $F < 1$. A direct comparison between Experiments 1 and 3 revealed a significant interaction between target quadrant and experiment in the mobile phase, $F(1, 30) = 19.29$, $p < 0.001$, $\eta_p^2 = 0.47$, suggesting that the lack of consistent viewer perspective of Experiment 1 was disruptive. Thus, locomotion did not interfere with acquiring probability cuing if the rich quadrant was stable relative to both the viewer and the monitor.

Experiment 4. Intentional learning

Explicit instructions to prioritize the rich quadrant of the screen led to a substantial performance gain in that quadrant. Search was more accurate when the target was in the rich quadrant (97.2%, $SE = 0.6\%$) than the sparse quadrants (96.3%, $SE = 0.7\%$), $t(15) = 2.90$, $p < 0.02$, Cohen's $d = 1.50$. As shown in Figure 4B, RT was significantly faster in the rich quadrant than the sparse quadrants in the mobile phase (Blocks 1–16), $F(1, 15) = 36.47$, $p < 0.001$, $\eta_p^2 = 0.71$. This effect was present as soon as the experiment started and did not further increase with training, $F(15, 225) = 1.41$, $p > 0.14$ for the interaction between quadrant condition

Experiment	Identified the rich quadrant			Failed to identify the rich quadrant		
	<i>N</i>	Rich	Sparse	<i>N</i>	Rich	Sparse
1	3	1949	2001	13	1963	2004
2	2	2445	2459	14	2410	2423
3	3	1602	1874	13	1805	2140
4	10	1838	2160	6	1578	1782

Table 1. Search RT (ms) during the mobile phase for participants who made different recognition choices.

and block. Participants were able to use the instructions to facilitate search in the rich quadrant as early as Block 1, $t(15) = 3.18$, $p < 0.006$, Cohen's $d = 1.64$. Thus, even though the rich quadrant was variable relative to the viewer's perspective, participants were able to use explicit instructions to prioritize a region of space in the external world. They may have done so by coding the rich quadrant according to either an environment-centered reference frame, or an egocentric reference frame that was updated as they moved. Although we cannot distinguish between these two possibilities, explicit instructions changed the pattern of performance compared with incidental learning. A direct comparison between Experiments 1 and 4 revealed a significant interaction between quadrant condition and experiment in the mobile phase, $F(1, 30) = 17.73$, $p < 0.001$, $\eta_p^2 = 0.37$.

Probability cuing disappeared in the testing phase when participants were informed of the target's (now random) distribution. An ANOVA on quadrant condition (previously rich or sparse) and block (17–20) revealed no effects of quadrant condition, $F < 1$, block, $F < 1$, or their interaction, $F(3, 45) = 1.28$, $p > 0.25$.

Recognition

The proportion of participants correctly identifying the high-frequency quadrant was 18.8% in Experiment 1, 12.5% in Experiment 2, and 18.8% in Experiment 3, all of which were no higher than chance (25%). All participants in Experiment 4 reported that they had followed the instruction to prioritize the rich quadrant (the two additional participants who ignored the instructions were removed, see Methods). However, not everyone believed that the instructions were valid. Only 62.5% of the participants in Experiment 4 chose the instructed quadrant as the quadrant where the target was most often found. To examine whether recognition correlated with search performance, we separated participants based on whether they were correct in the recognition test. Table 1 shows search RT in the mobile phase for the rich and sparse conditions. In no experiments did recognition performance interact with quadrant condition, all p values > 0.20 .

Discussion

This study characterized visual search behavior in participants who viewed the display from variable perspectives. In these experiments one quadrant was more likely to contain the target than the other quadrants, and the target-rich quadrant was always fixed in the external environment (e.g., on the monitor). In three experiments learning was incidental. In a fourth experiment participants were encouraged to prioritize the target-rich quadrant. We found that participants were able to intentionally prioritize the rich quadrant with explicit instructions. Moving around did not prevent them from prioritizing a rich quadrant that was fixed in the environment but variable relative to themselves. The finding is consistent with one of two possibilities—participants may have coded the target-rich quadrant in an environment-centered reference frame (e.g., the landmarks in the environment), or they may have coded this region in a viewer-centered reference frame but successfully performed spatial updating as they moved around. Regardless of which possibility was true, participants could prioritize an environment-rich region. Thus, goal-driven atten-

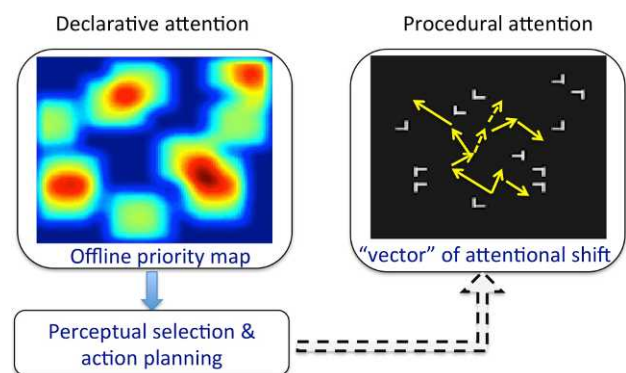


Figure 5. An illustration of the dual-system view. Spatial attention has a declarative and a procedural component. The declarative component specifies which locations to attend before the actual shift of attention. The output of the declarative attention feeds into perceptual selection and action planning. Procedural attention refers to the actual vector of attentional shift during attentional movement, which is represented here as arrows.

tion is relatively flexible in its reference to the external world and can serve as an important mechanism for spatial navigation and foraging. However, the intentional learning experiment revealed little evidence of *learning*. Participants were able to use the instructions immediately and did not more efficiently search the target-rich quadrant with additional training.

Results from intentional learning can be contrasted with those of incidental learning. In Experiments 1–3 participants received no explicit knowledge about where the target was likely to be. The first two experiments showed a lack of learning. Participants were unable to prioritize the target-rich quadrant based on incidental learning alone, even when search was conducted against a natural scene. This was not because locomotion itself interfered with learning. In the third experiment participants moved between trials but always returned to the same home base. Under these conditions probability cuing toward the rich quadrant was acquired. Like the other two experiments participants in Experiment 3 were not aware of the experimental manipulation. Explicit awareness cannot account for the difference in search. Instead, when the viewer-centered reference frame was consistently aligned with the environment-centered reference frame (Experiment 3), the target-rich regions could be coded relative to the participants' body and the external environment. This situation yielded learning. The comparison across the three incidental-learning experiments suggests that incidentally learned attention is egocentric, and that spatial updating does not adequately compensate for its disruption following locomotion. When search was conducted from random viewpoints probability cuing did not occur (Experiments 1 and 2). The lack of learning was observed despite almost 400 trials of training, far more than the amount necessary to acquire probability cuing in Experiment 3. These data provide compelling evidence for the hypothesis that frequently attended locations, when acquired incidentally, are coded egocentrically and are not adequately updated.

Together these experiments provide strong evidence for the existence of two dissociable systems of spatial attention. In previous research, spatial attention was often conceptualized as an activation map that prioritizes salient and behaviorally relevant stimuli (Fecteau & Munoz, 2006; Itti & Koch, 2001). The priority map is highly sensitive to top-down knowledge, which includes both explicit goals and implicit learning (Chun, 2000; Wolfe, 2007). In contrast, our study suggests that it is important to distinguish between the two sources of top-down knowledge. Goal-driven attention is flexible and can be directed to a region of the environment independent of the viewer's perspective. Incidental learning depends on the stability of frequently attended locations relative to the viewer.

Top-down, goal-driven attention and incidentally learned attention can be dissociated in several other ways. First, they differ in their flexibility and persistence. Goal-driven attention can be directed from trial-to-trial to different parts of the display (Jiang, Swallow, & Rosenbaum, 2013). Incidentally learned attention shows persistence and high resistance to extinction (Jiang, Swallow, Rosenbaum, & Herzig, 2013). Second, the two types of attention differ in their time course. Goal-driven attention occurs shortly after the spatial cue, typically 100–300 ms later (Posner, 1980; Vickery, King, & Jiang, 2005). The cue is used to deploy attention before the presentation of the display. This is not the case with incidentally learned attention. For example, in contextual cuing, the repeated spatial context does not *begin* to cue attention until search is underway (Jiang, Sigstad, & Swallow, 2013).

We can characterize incidental learning as a form of “procedural attention” and goal-driven attention as a form of “declarative attention.” The former refers to the “online” shift of attention. Each shift that brings one closer to the target is reinforced and is more likely to occur in the future. The latter corresponds to the activation map that specifies the attentional priority of locations in space. The priority map feeds into perceptual selection and action planning, but its impact on attentional shifts is “offline.” Figure 5 illustrates the *dual-system view* of attention.

By using the terms “declarative” and “procedural,” we suggest that the division of spatial attention resembles the division of human memory (Schacter, 1996; Squire, 1992, 2004). However, although the implicit learning literature has long debated whether implicit and explicit learning are a single system (Shanks, 2005) or two systems (Sanchez & Reber, 2013), few implicit learning paradigms tap into spatial attention. To the extent that spatial locations (rather than sequences of visuomotor actions) are learned, learning appears to be explicit rather than implicit (Witt & Willingham, 2006). In fact, we believe that the underlying brain mechanisms for declarative attention are largely distinct from those for declarative memory (Duncan, 2010; Squire, 2004), while there may be some but not full overlap between those for procedural attention and procedural memory (Chun & Phelps, 1999; Graybiel, 2008).

The dual-system view is reminiscent of the premotor theory of attention (Rizzolatti, Riggio, Dascola, & Umiltà, 1987), or the idea that attention is for visuomotor action (Allport, 1989; Tipper et al., 1998). However, it can be distinguished from these other theories. Unlike the premotor theory, we believe that procedural attention is only one component of attention. In addition, although existing theories of attention acknowledge that “shift” is a component of attention (e.g., Posner & Petersen, 1990), they do not distinguish between shifts initiated by top-down goals

and shifts influenced by implicit learning. The dual-system view proposes that whereas declarative attention feeds into action planning, its influence occurs prior to an attentional shift—it determines where attention is likely to go. Procedural attention, on the other hand, is involved online, during the actual process of moving attention in space.

A long-standing tradition in attention research is to classify attention into multiple systems (Awh, Belopolsky, & Theeuwes, 2012; Chun, Golomb, & Turk-Browne, 2011; Desimone & Duncan, 1995; Egeth & Yantis, 1997; Pashler, 1994). Theories have emphasized the distinction between perceptual attention and central attention (Pashler, 1994; to some degree Chun et al., 2011), between top-down and bottom-up attention (Desimone & Duncan, 1995; Egeth & Yantis, 1997; Wolfe, 2007), and between history, reward-driven, and more immediate forms of attention (Anderson, Laurent, & Yantis, 2011; Awh et al., 2012). The dual-system view builds upon existing theories but also differs from them. Goals and saliency (and to some degree reward history) both modulate the weight of attentional priority (Figure 5). The output of the priority map interfaces with perceptual selectivity and action planning (Fecteau & Munoz, 2006). Declarative attention, much like Milner and Goodale's (2008) ventral system, is an offline system. It makes the plan but does not actually make the shift. Separate from these influences is a form of procedural attention, represented as “vectors” of attentional shift acquired in the actual process of the task. Procedural attention executes the plan. Each attentional shift leads one to another location, and the vector that results in the detection of the target is reinforced. In the case of probability cuing, the vector of attention is systematically biased toward the target-rich regions, and hence it is reinforced. Critically, the vector is coded relative to the viewer, so when the viewpoint is variable, no consistent vectors are reinforced even when the target-rich regions are fixed on the display.

Probability cuing affects spatial attention rather than oculomotor responses because eye movement toward the target-rich locations is neither sufficient nor necessary for probability cuing. Probability cuing occurs when eye movement is prevented (Geng & Behrmann, 2005), and when the display is presented so briefly that there is no time to move one's eyes (Jiang & Swallow, 2013). In addition, frequently moving one's eyes toward certain locations is insufficient to yield probability cuing (Jiang, Swallow, & Rosenbaum, 2013). Thus, probability cuing is unlikely an oculomotor routine. More importantly, probability cuing fulfills the definition of attentional guidance (Wolfe, 1994, 2007). Visual search slope is reduced when the target appears in the rich quadrant than the sparse quadrants, and this reduction is comparable to that induced by a

central arrow (Jiang, Swallow, & Rosenbaum, 2013). Thus, probability cuing is a form of attention but also differs qualitatively from goal-driven attention.

To return to the problem of spatial navigation and foraging, we believe that humans have at least two systems of attention. The explicit, intentional system flexibly prioritizes a region of the environment that may not be consistent relative to the viewer. In addition, humans have a system based on implicit knowledge that codes attended locations relative to their body. The latter system is helpful in some circumstances. Humans do not approach a destination from completely random directions. Navigation paths are usually constrained by environmental structures, allowing an egocentric system to benefit performance. In addition, the egocentric system is useful because it can naturally relate to the visuomotor system, which codes objects in a viewer-centered reference frame (Goodale & Haffenden, 1998; Milner & Goodale, 2008). The combination of goal-driven attention and incidentally learned attention can solve the problems of real-world foraging and navigation.

This study does not indicate which egocentric system is used to code incidentally learned attention. It may be head centered, eye centered, or body centered. This study also does not distinguish between an environment-centered representation and an egocentric-but-updated representation for goal-driven attention. In addition, although our classification of the spatial reference systems is similar in complexity to other studies on attention (e.g., spatiotopic vs. retinotopic attention), it is simpler than the classifications used in spatial memory research, which distinguishes viewpoint dependence from egocentric and allocentric representations (e.g., Mou, Fan, McNamara, & Owen, 2008; Wang, 2012). We have simplified our classification here because we focus on the representation of one region (the rich quadrant) rather than multiple locations, which are typically the focus of spatial memory research. Despite this simplification, our data make it clear that moving around during search interferes with the ability to acquire environmental regularities.

Finally, although we have shown that consistent viewpoint is necessary to acquire implicit probability cuing, we do not know whether this condition is *sufficient*. It is possible that consistent viewpoint and consistent environment are jointly necessary for incidentally learned attention. These questions should be addressed in future research.

Conclusions

By testing observers who move to different perspectives for each trial of visual search, this study has

revealed two mechanisms that code frequently attended locations. Whereas goal-driven attention can be used to prioritize a region of the environment independent of the viewer's perspective, consistent viewer-centered representations are necessary for incidentally learned attention. Our study is consistent with the claim that attention can be divided into two systems—declarative attention and procedural attention. In addition, our finding constrains the ubiquity of visual statistical learning. When learning occurs incidentally in visual search, environmental statistics are useful only when they can be effectively coded in a viewer-centered reference frame. The University of Minnesota funded this work.

Keywords: attention, visual search, incidental learning, spatial reference frame

Acknowledgments

We thank Steve Engel, Joy Geng, Wilma Koutstaal, and Bryce Palm for comments on an earlier draft, and Gail Rosenbaum for data collection. Author contribution is as follows. YVJ and KMS developed the study concept. All authors contributed to the design. YVJ and CGC set up the experiments and collected the data. YVJ and KMS interpreted the data and wrote the paper. All authors approved the final version of the paper for submission.

Commercial relationships: none.

Corresponding author: Yuhong Jiang.

Email: jiang166@umn.edu.

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

References

- 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.
- Allport, D. A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631–682). Cambridge, MA: MIT Press.
- Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*, 10367–10371.
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Science*, *16*, 437–443.
- Ball, K., Smith, D., Ellison, A., & Schenk, T. (2010). A body-centered frame of reference drives spatial priming in visual search. *Experimental Brain Research*, *204*, 585–594.
- Ball, K., Smith, D., Ellison, A., & Schenk, T. (2009). Both egocentric and allocentric cues support spatial priming in visual search. *Neuropsychologia*, *47*, 1585–1591.
- Behrmann, M., & Tipper, S. P. (1999). Attention accesses multiple reference frames: Evidence from visual neglect. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 83–101.
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*, 77–80.
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.
- Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception & Performance*, *33*, 798–815.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
- 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.
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*, 170–178.
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.
- Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 224–234.
- Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, *2*, 844–847.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelli-

- gent behavior. *Trends in Cognitive Sciences*, *14*, 172–179.
- Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.
- Fecteau, J. H., & Munoz, D. P. (2006). Saliency, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*, 382–390.
- Geng, J. J., & Behrmann, M. (2002). Probability cuing of target location facilitates visual search implicitly in normal participants with hemispatial neglect. *Psychological Science*, *13*, 520–525.
- Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception & Psychophysics*, *67*, 1252–1268.
- 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.
- Golomb, J. D., Pulido, V. Z., Albrecht, A. R., Chun, M. M., & Mazer, J. A. (2010). Robustness of the retinotopic attentional trace after eye movements. *Journal of Vision*, *10*(3):19, 1–12, <http://www.journalofvision.org/content/10/3/19>, doi:10.1167/10.3.19. [PubMed] [Article]
- Goodale, M. A., & Haffenden, A. (1998). Frames of reference for perception and action in the human visual system. *Neuroscience & Biobehavioral Reviews*, *22*, 161–172.
- Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, *31*, 359–387.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203.
- Jiang, Y. V., Sigstad, H. M., & Swallow, K. M. (2013). The time course of attentional deployment in contextual cueing. *Psychonomic Bulletin & Review*, *20*, 282–288.
- Jiang, Y. V., & Swallow, K. M. (2013). Spatial reference frame of incidentally learned attention. *Cognition*, *126*, 378–390.
- Jiang, Y. V., Swallow, K. M., & Rosenbaum, G. M. (2013). Guidance of spatial attention by incidental learning and endogenous cuing. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 285–297.
- 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.
- Jiang, Y. V., Swallow, K. M., & Sun, L. (2013). Egocentric coding of space for incidentally learned attention: Effects of scene context and task instructions. Manuscript submitted for publication.
- Johnson, J. S., Woodman, G. F., Braun, E., & Luck, S. J. (2007). Implicit memory influences the allocation of attention in visual cortex. *Psychonomic Bulletin and Review*, *14*, 834–839.
- Mou, W., Fan, Y., McNamara, T. P., & Owen, C. B. (2008). Intrinsic frames of reference and egocentric viewpoints in scene recognition. *Cognition*, *106*, 750–769.
- Mathot, S., & Theeuwes, J. (2010). Gradual remapping results in early retinotopic and late spatiotopic inhibition of return. *Psychological Science*, *21*, 1793–1798.
- Miller, J. (1988). Components of the location probability effect in visual search tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 453–471.
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*, 774–785.
- Nakamura, K., & Colby, C. L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Sciences*, *99*, 4026–4031.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*, 220–244.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pertsov, Y., Zohary, E., & Avidan, G. (2010). Rapid formation of spatiotopic representations as revealed by inhibition of return. *Journal of Neuroscience*, *9*, 1–12.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bonwhuis (Eds.), *Attention & performance X: Control of language processes* (pp. 551–556). Hillsdale, NJ: Erlbaum.
- Posner, M. I., & Peterson, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.
- Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 1157–1165.

- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*, 31–40.
- Rock, I. (1985). *The logic of perception*. Cambridge, MA: MIT Press.
- Sanchez, D. J., & Reber, P. J. (2013). Explicit pre-training instruction does not improve perceptual-motor sequence learning. *Cognition*, *126*, 341–351.
- Schacter, D. J. (1996). *Searching for memory: The brain, the mind, and the past*. New York: Basic Books.
- Shanks, D. (2005). Implicit learning. In K. Lamberts & R. L. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220), London: Sage Publications.
- Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science*, *9*, 315–320.
- Squire, L. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*, 195–231.
- Squire, L. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*, 171–177.
- Tipper, S. P., Howard, L. A., & Houghton, G. (1998). Action-based mechanisms of attention. *Philosophical Transactions of the Royal B: Biological Sciences*, *353*, 1385–1393.
- Treisman, A. (1988). Features and objects: The fourteenth Barlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, *40*, 201–237.
- Tsuchiai, T., Matsumiya, K., Kuriki, I., & Shioiri, S. (2012). Implicit learning of viewpoint-independent spatial layouts. *Frontiers in Psychology*, *3*, 207.
- Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1):8, 81–92, <http://www.journalofvision.org/content/5/1/8>, doi:10.1167/5.1.8. [PubMed] [Article]
- Wang, R. F. (2012). Theories of spatial representations and reference frames: What can configuration errors tell us? *Psychonomic Bulletin and Review*, *19*, 575–587.
- Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition*, *70*, 191–210.
- Witt, J. K., & Willingham, D. T. (2006). Evidence for separate representations for action and location in implicit motor sequencing. *Psychonomic Bulletin & Review*, *13*, 902–907.
- Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York: Oxford University Press.
- Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research*, *34*, 1187–1195.
- Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research*, *48*, 2070–2089.