Effects of attention without perceptual awareness on motor responses, memory, and behavior

Jeffrey Y. Lin

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington
2012

Program Authorized to Offer Degree:
Department of Psychology

University of Washington

Abstract

Effects of attention without perceptual awareness on motor responses, memory and behavior

Jeffrey Y. Lin

Chairs of the Supervisory Committee:

Associate Professor, Geoffrey M. Boynton
Associate Professor, Scott O. Murray
Professor, Geoffrey R. Loftus

Department of Psychology

Our everyday visual experience is strongly affected by attention. Visual attention can enhance or prioritize the processing of specific stimuli over the overwhelming number of other sensory inputs by selecting spatial locations, features, objects, and even time. For example, attending to a particular feature such as the color of an object produces a global facilitation of processing for stimuli that share that feature; alternatively, cueing attention to a particular location can enhance sensitivity to visual input at the cued location. A critical question for understanding the relationship between attention and consciousness is whether awareness is required for this type of prioritized attentional selection. It has been suggested that visual attention can only be affected by consciously perceived events; however, we identified three novel and surprising results about the nature of attention and how it can influence our motor responses, memory and behavior without perceptual awareness. (1) We demonstrated how the visual system can extract behaviorally relevant details from a visual scene and automatically categorize threatening versus non-threatening images at a level of precision beyond our conscious perceptual capabilities in the absence of perceptual awareness. (2) We found that memory for scenes was enhanced when presented concurrently with a behaviorally important target—this is evidence of a mechanism where traces of a visual scene are automatically encoded into memory at behaviorally relevant points in time regardless of the spatial focus of attention. (3) We found evidence for a previously unknown exogenous cueing mechanism for feature-based attention where visual attention responds reflexively and rapidly in response to color cues.

TABLE OF CONTENTS

Capture of attention to threatening stimuli without perceptual awareness

Jeffrey Y. Lin[1], Scott O. Murray[1], Geoffrey M. Boynton[1]
jytlin@u.washington.edu, somurray@u.washington.edu, gboynton@u.washington.edu

Department of Psychology, University of Washington[1]

Corresponding Author:
Jeffrey Y. Lin [jytlin@u.washington.edu]

Vision and Cognition Group
Department of Psychology
University of Washington
4000 15th Ave NE
Seattle, WA 98195-1525
United States
206.293.9359

Introduction

Visual images that convey threatening information can automatically capture attention [1-4]. One example is an object looming in the direction of the observer—presumably because such a stimulus signals an impending collision [5]. A critical question for understanding the relationship between attention and conscious awareness is whether awareness is required for this type of prioritized attentional selection [6]. Although it has been suggested that visual spatial attention can only be affected by consciously perceived events [7], we show that automatic allocation of attention can occur even without conscious awareness of impending threat. We used a visual search task to show that a looming stimulus on a collision path with an observer captures attention but a looming stimulus on a near-miss path does not. Critically, observers were unaware of any difference between collision and near-miss stimuli even when explicitly asked to discriminate between them in separate experiments. These results counter traditional salience-based models of attentional capture, demonstrating that in the absence of perceptual awareness, the visual system can extract behaviorally relevant details from a visual scene and automatically categorize threatening versus non-threatening images at a level of precision beyond our conscious perceptual capabilities.

Methods

Participants. 12 undergraduates at the University of Washington (8 females, 4 males) received financial compensation for participating in Experiment 1. 20 undergraduates (12 females, 8 males) received financial compensation for participating in either Experiment 2A or 2B. All reported normal or corrected-to-normal visual acuity and maintained an overall accuracy better than 80%. All subjects gave informed consent to participate in this experiment, which was approved by the University of Washington Human Subjects Institutional Review Board.

Apparatus, stimuli and procedure. Displays were generated in Matlab (Mathworks) using the psychophysics toolbox [35, 36] and were presented on a 51-cm (diagonal) Samsung Syncmaster 1100DF CRT monitor at 1024x768 resolution, refreshed at 60 Hz in a room with no ambient lighting. Participants used a chinrest and sat with their eyes 50 cm from the screen. The background of the displays was gray (15 cd/m^2). Display items consisted of discs (4.6 degrees of visual angle) filled with a linear shading gradient that ran from white (30 cd/m^2) in the top right to black (0.1 cd/m^2) in the bottom left, giving the impression that they were spheres lit from above and to the right.

A video clip of a typical trial is available (see Appendix C). Each trial in Experiment 1 consisted of four stages. (1) The initial preview display lasted 33 ms and consisted of a small fixation dot and a display of three or six spheres in the locations that would be used for the final search display. (2) This was followed by a looming display where a sphere expanded uniformly from a small size (2.1 deg) to the standard sphere size (4.6 deg) in 125 ms across 7 frames of motion (60 frames/sec) towards one of eight locations around the boundaries of the observer's head. Looming stimuli were defined in 3D real-world coordinates and rendered on a monitor 50 cm from the observer using perspective geometry. Looming stimuli were spheres 8 cm in diameter that moved from a distance of 350 cm from the observer to 175 cm from the observer over 125 ms; this motion corresponded with a degree change of 2.1 to 4.6 degrees of visual angle. Looming stimuli that represented a collision with the observer's head had an initial position 6 cm from the center of the monitor (6.9 deg) and the trajectory simulated a point of impact 3 cm from the center of the observer's head. Looming stimuli that represented a near-miss had an initial position 5 cm from the center of the monitor

(5.7 deg) and simulated a final impact point 6 cm from the center of the observer's head. At the end of the looming animation, both collision and near-miss stimuli had identical end points 8 cm from the center of the monitor (9.1 deg). Half of the trials displayed a looming item on a path representing a collision with the observer, and the other half of the trials displayed a looming item on a path that represented a near-miss with the observer. The eight final locations of the spheres were positioned evenly around fixation with radii of 9.1 deg (clock positions: 12, 1:30, 3, 4:30, 6, 7:30, 9, and 10:30). (3) This looming motion display was followed by a 16 ms blank screen which was inserted before the presentation of the search display to mask the local pop-out deformation that occurred when the target item transformed from a sphere into an oval. (4) The blank screen was followed by the search display, which remained in view until participants responded or 2000 ms elapsed. In all search displays, a target oval was created by narrowing spheres by 5.5% (from 4.6 to 4.3 deg) along either the horizontal or vertical dimension. Different conditions were counterbalanced and randomized in every block for every participant.

Participants were instructed to search for the oval (while maintaining fixation at the center fixation point) and to discriminate its orientation (vertical or horizontal) as rapidly as possible by pressing one of two keys. A small plus sign (correct), minus sign (incorrect), or circle (no response) provided feedback, and was replaced by a circle to serve as the new fixation point and signal the start of the next display. Participants were informed that every display would have a looming item and a target oval, but that the final location of the looming item provided no information about the location of the target oval. The locations of the looming item and target oval were determined randomly and independently in every trial such that the looming item and the target oval coincided at the same location every 1/n trials (with "n" being the display size of either three or six). Participants were instructed to respond as quickly as possible while maintaining an accuracy of at least 80%. Prior to testing, participants received 54 practice trials. Each participant was tested for a total of 540 trials, in 5 blocks of 108 trials. Blocks were separated by brief breaks.

For Experiment 2A, looming stimuli from the 6 o’clock position were removed due to the fact that all looming stimuli from this location were classified as collisions in pilot testing; displays were otherwise identical to those in Experiment 1. Participants were told that each display contained one of two types of looming items—collision looms that would travel a path towards a collision with their head or body and miss looms that would travel a wider path and miss their head or body. Participants were then instructed to fixate at the center, attend to the looming items and report which type of looming stimulus was displayed in the trial. Conditions were randomized in every block and feedback was not provided in this version of the experiment. Prior to testing, participants received 54 practice trials, and each participant was tested for a total of 216 trials, in two blocks of 108 trials.

For Experiment 2B, the displays were identical to those used in Experiment 1. Critically, two trajectories clearly representing a collision and a miss were added to the displays and randomized with the subtle trajectories used in Experiment 1. Collision stimuli in Experiment 2B simulated points of impact of 1 cm and 3 cm from the center of the observer’s head while miss stimuli simulated points of impact of 6 cm and 12 cm from the center of the observer’s head. In addition, feedback was provided in this experiment. The task was identical to the task in Experiment 2A. Prior to testing, participants received 54 practice trials. Each participant was tested for twice the number of trials as the participants in Experiment 2A for a total of 432 trials, in 2 blocks of 216 trials.

Experiment 1 Results

Each trial began with a looming stimulus followed by a search display where participants were instructed to quickly locate and discriminate the orientation of a target oval amongst a field of distracting circular discs. Targets and distractors were placed in eight possible positions in a circular array around the point of fixation. Trials varied by (1) *path:* whether or not the looming stimulus was on a collision path with the boundary of the observer's head, (2) *position:* whether or not the final position of the looming stimulus coincided with the target oval (versus a distractor disc), and (3) *display size:* the size of the search array (either 3 or 6 items). Looming stimuli on collision and near-miss paths had the same final positions. This paradigm allowed us to measure the effects of the path of looming stimuli on search rates when their final positions either did or did not coincide with the spatial location of the target oval.

Figure 1A shows mean response times (RT) for trials on which participants correctly determined the orientation of the target oval. Figure 1B shows mean error rates. Individual results for each subject are also available (see Appendix A). Increased search efficiency for *collision targets* (targets at locations that followed looming stimuli on a collision path) was evident in the absolute search times for a set size of six items (128.8 ms difference between collision targets and near-miss targets). Increased search efficiency was also evident in the rate of search as indexed by search slopes: rates of search were fastest for collision targets (-1.7 ms/item), and were much slower for near-miss targets (46.1 ms/item), collision distractors (42.4 ms/item) and near-miss distractors (57.8 ms/item).

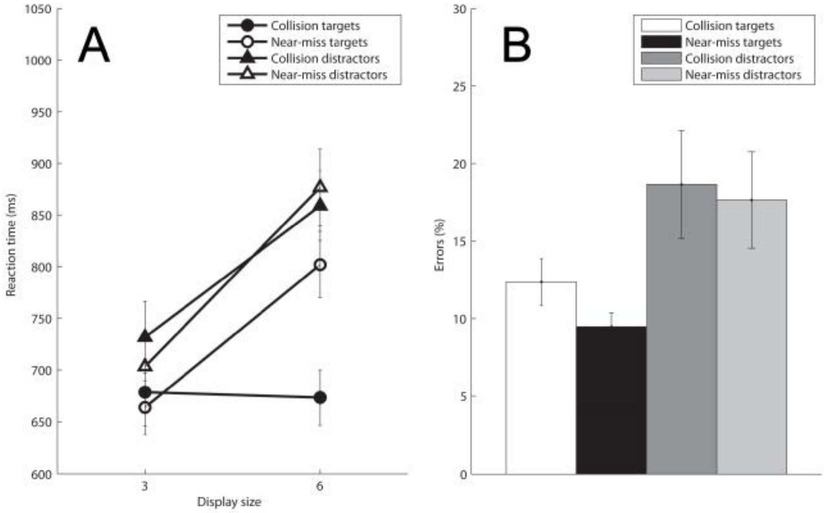

Figure 1 – Mean correct response times and error rates in Experiment 1
(A) Collision targets and near-miss targets represent trials where the final position of the looming items coincided with the location of the target ovals. Collision distractors and near-miss distractors represent trials where the final position of the looming item was at a location away from the location of the target oval. Increased search efficiency was evident in the rate of search as indexed by search slopes: rates of search were fastest for collision targets (-1.7 ms/item), and were much slower for near-miss targets (46.1 ms/item), collision distractors (42.4 ms/item) and near-miss distractors (57.8 ms/item). A significant three-way interaction of position, path, and display size (F(2,7) = 7.56, p = 0.019, MSE = 2005.71) indicates significant differences in the slopes across the four path/position combinations. Error bars represent s.e.m.
(B) Error rates for the different conditions of Experiment 1 are presented in Figure 1B. Error bars represent s.e.m.

A repeated-measures ANOVA on RTs from correct trials indicated main effects of looming position (RTs were faster for collision targets than collision distractors, F(1,11) = 29.7, $p < 0.0001$, MSE = 8080.17), path (RTs were faster for collision targets than near-miss targets, F(1,11) = 8.68, p = 0.013, MSE = 1658.68), and display size (RTs were smaller for set size three than for set size six, F(1,11) = 60.7, $p < 0.0001$, MSE = 3852.03). A significant three-way interaction of position, path, and display size (F(2,7) = 7.56, p = 0.019, MSE =

2005.71) indicates significant differences in the slopes across the four path/position combinations.

These results show that the location of a looming stimulus that was on a collision course with the subject's head received prioritized attention in the visual search process that followed. Due to the nature of the paradigm, all looming items that originated in the 6 o' clock (downward) location traveled on a path upward towards the observer's body. Figure 2 shows that response times for collision targets at the 6 o'clock position (595.01 ms) were significantly faster than response times for collision targets towards the head (675.94 ms), $t(33) = -3.08$, $p < 0.0001$.

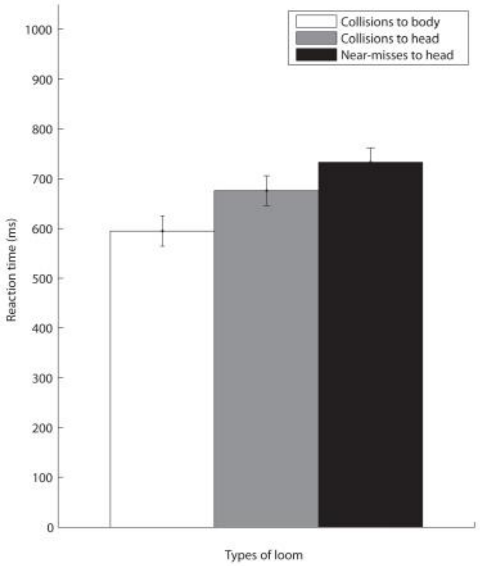

Figure 2 – Mean correct response times for collisions to the body versus the collisions to the head in Experiment 1
Body collisions and head collisions represent trials where looming items coincided as the target ovals; body collisions represented looming items that originated from the 6 o' clock location and traveled towards the observer's body, while head collisions represented looming items in every other location that traveled toward the observer's head. Reaction times to body collisions (595.01 ms) were significantly faster than reaction times to head collisions (675.84 ms), $t(33) = -3.08$, $p < 0.0001$. Error bars represent s.e.m.

When briefly questioned after the experiment, all participants reported being subjectively unaware of any differences in the trajectories of looming stimuli. Surprisingly, most subjects also reported being able to ignore the looming stimuli since they provided no information about the target detection task and were only present over 125 ms.

Experiment 2 Results

Two versions of a control experiment were conducted to directly test subjects' ability to discriminate between the two looming paths used in Experiment 1. In Experiment 2A the 6 o'clock position was removed because all looming items originating from this location traveled on paths towards the observer's body and were easily identified during pilot testing. The display was otherwise identical to Experiment 1. Participants were instructed to attend to the looming item in each display while fixating at the center and discriminate the trajectory of the looming stimulus as either a collision or a near-miss with their head or body. Figure 3A shows mean accuracy (50.33 +/- 1.41%) and mean sensitivity (d' = 0.171 +/- 0.180) for discriminating the trajectory of the looming item in each trial. No feedback was provided. Figure S2 shows individual mean accuracies and sensitivity measures (see Appendix B).

In Experiment 2B, there were four critical additions: (1) we re-introduced the 6 o'clock position, making the displays identical to the displays used in Experiment 1, (2) we implemented two additional trajectories representing a clear miss and clear collision trajectory and randomized them with the subtly different trajectories used in Experiment 1 for a total of four trajectories in the experiment, (3) we added feedback to every trial, and (4) we doubled the number of trials each participant conducted. Given every opportunity to learn this task, the results were surprising. Figure 3B shows mean accuracy (86.54 +/- 4.57%) for discriminating between the "clear trajectories" and mean accuracy (53.82 +/- 3.70%) for discriminating between the "subtle trajectories" used in Experiment 1. These results suggest that subjects understood the task and could easily discriminate clear collisions from clear miss trajectories; however, participants were unable to accurately classify the subtly different trajectories presented in Experiment 1.

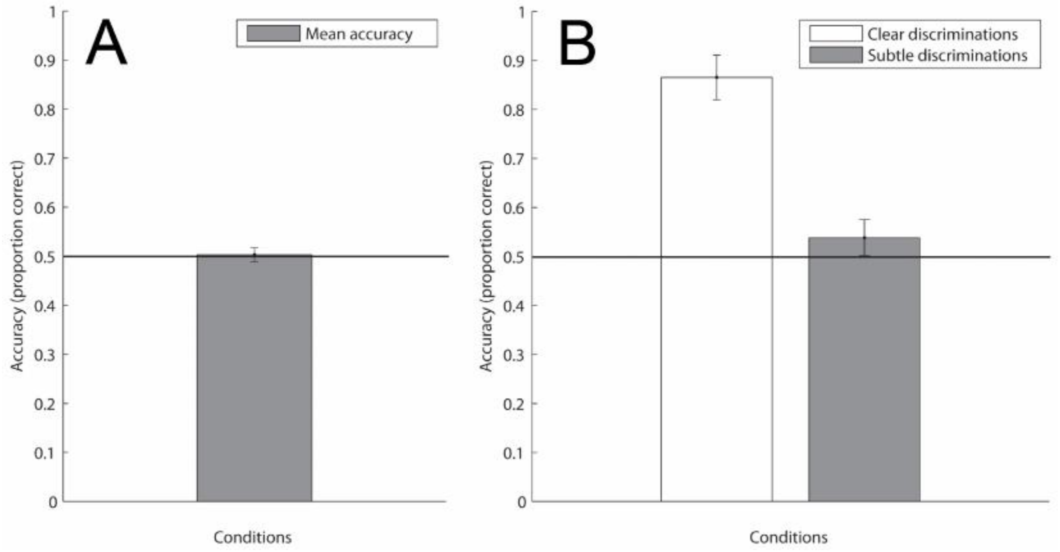

Figure 3 – Mean accuracies for Experiment 2
(A) Mean accuracy (50.33 +/- 1.41%) and mean sensitivity (d' = 0.171 +/- 0.180) for discriminating the trajectory of the subtle looming items are plotted from the results of Experiment 2A. Participants were instructed to fixate at the center, attend to the looming items and identify whether the looming trajectory was closer to a collision or near-miss with their head. Error bars represent s.e.m ($p = 0.001$).
(B) Mean accuracy (86.54 +/- 4.57%) for discriminating between the "clear trajectories" and mean accuracy (53.82 +/- 3.70%) for discriminating between the "subtle trajectories" used in Experiment 1 are plotted from the results of Experiment 2B. Error bars represent s.e.m. Results suggest that participants understood the task and could discriminate clear collision from miss trajectories; however, participants were unable to categorize the subtle looming trajectories presented in Experiment 1, $p < 0.05$.

Discussion

Attentional capture can be operationally defined as speeded search performance that is independent of set size when a nonpredictive stimulus happens to be at the target location. A classic example is a visual onset that is searched with priority, even when it is irrelevant to the main task [8, 9]. Reaction time for detecting a target plotted as a function of the number of distractors can be used as an index for attentional capture: flat search slopes indicate attentional capture and steep slopes a failure to capture attention.

Perceptual saliency is often considered to be a primary factor in determining whether or not a target captures attention [10-12]. Typically, a saliency map is calculated by assigning each visual location a saliency value obtained by the summation of activation values from separate feature maps [13-16]. Indeed, stimulus-driven perceptual saliency models with maps for features such as color, contrast and motion can account for a wide variety of behavioral effects observed in search tasks [17]. However, not all attention-capturing differences between stimuli can be described by saliency models. For example, visual stimuli that convey threatening information can capture attention because of their obvious behavioral relevance, but may share similar features with non-threatening stimuli. In fact, the appearance and behavior of predators in natural environments has typically evolved to minimize visual salience.

To maximize survival, threatening information should quickly and automatically capture attention even when this information is perceptually non-salient. However, traditional perspectives on attentional selection have also suggested that for an event to capture attention, the event needs to be consciously perceived by the observer. For example, there are numerous cases of exogenous cues that enhance sensitivity to visual input at a cued location [18, 19]. We reasoned that not only should threatening information automatically prioritize attention, but may even rely on separate, unique neural processes that are independent of conscious perception [20-23].

Recently, it has been shown that an otherwise uninformative motion stimulus at the target location can capture attention provided it is on a collision path with the observer [3]; however, the two types of looming stimuli used in this previous study traveled in completely opposite directions – either towards fixation or away from fixation – and were thus easily distinguishable. Here, we report a dissociation between attention and awareness in which a looming object on a collision path with the observer captures attention (Experiment 1) even though it is perceptually indistinguishable from a looming stimulus that just misses the observer (Experiment 2). It should be noted that attentional priority was given to a looming stimulus on a collision path only when the looming stimulus coincided with the target oval. Collision or near-miss distractors were equally effective in drawing attention away from stationary target ovals and delaying reaction times. This means that the attentional effects reported here were very specific in spatial location and not a general effect of arousal due to the presence of a threatening stimulus. This spatial specificity shows that this automatic attentional process has spatially-selective detectors and is therefore more sophisticated than a simple general threat-detecting mechanism.

Surprisingly, looming targets from the 6 o' clock position that traveled towards the observer's body produced the fastest response times relative to looming targets that traveled towards the observer's head. This lends support to the idea that attentional capture without awareness is not a binary process, but rather can vary continuously in strength [24]. For some reason, looming stimuli from the 6 o'clock position appears to have the most threatening direction of motion. This is consistent with the idea that relative threat is assessed from a center-of-mass model. It is important to note that when using a manual response paradigm to study the attentional effects of behaviorally relevant stimuli, there is always the possibility of eye saccade influences upon the manual responses. For example, some studies have shown that threatening stimuli only capture attention when presented for 50 ms when using eye saccade measurements, while threatening stimuli only capture attention when presented for 500 ms when using manual responses [25].

Our results have two significant implications for models of visual processing and attention. First, the results extend recent empirical demonstrations of a dissociation between attention and awareness. There are at least three ways in which attention might operate without awareness. First, cues may influence attention by virtue of contingencies between the cue and target for which the subject was unaware. For example, Bartolomeo et al. (2007) demonstrated increased performance in a target detection task utilizing valid cues even though subjects were not explicitly aware that the cues were valid. Second, a target may be processed more effectively by virtue of being attended while the subject remains unaware of it. For example, research has shown that attending to color cues that were rendered invisible through metacontrast masking resulted in enhanced discrimination in a subsequent color discrimination task. [26]. Third, attention may be directed by cues the subject is unaware of. Structurally, such experiments are similar to traditional attention cueing paradigms except that the attentional cues are presented in such a way that subjects are unaware of their presence. For example, Jiang et al. (2006) used inter-ocular suppression to show that “invisible” erotic images that presumably never reach conscious awareness managed to repel or attract attention. Zhaoping (2008) used inter-ocular suppression to show that an eye-of-origin or ocular singleton can attract attention even though observers were unaware that an item was presented to the left eye among background items presented to the right eye. Notably, these previous experiments have relied on rather unnatural stimulus masking manipulations to render the cues invisible. Though our experiment also falls under the category of using a cue the subject is unaware of, our stimuli are unique because our motion cue itself is fully visible. Even though subjects are fully aware of the presence of motion stimuli, differences below discrimination thresholds still produce differential effects on detection.

A second major implication of our results is that they support the influential theory of visual processing that suggests two independent systems within the visual system; one supporting conscious perception while the other unconsciously guiding our actions [27-29]. Evidence for this has typically involved illusions that affect perception but not the sensorimotor systems [30-33] or involved special populations such as patients with posterior cerebral lesions [34]. Intuitively, reacting to threatening stimuli, such as a predator attack, should not require the time-consuming process of consciously identifying the species or identity of the predator. The present study shows that indeed the unconscious action of directing attention to the location of a potentially threatening looming stimulus appears to be automatic and unconscious and is, surprisingly, more accurate at calculating an object's path of motion than the conscious perception pathway.

REFERENCES

1. Ball, W., and Tronick, E. (1971). Infant responses to impending collision: optical and real. Science *171*, 818-820.

2. Koster, E.H., Crombez, G., Van Damme, S., Verschuere, B., and De Houwer, J. (2004). Does imminent threat capture and hold attention? Emotion *4*, 312-317.

3. Lin, J.Y., Franconeri, S., and Enns, J.T. (2008). Objects on a collision path with the observer demand attention. Psychol Sci *19*, 686-692.

4. Schiff, W., Caviness, J.A., and Gibson, J.J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". Science *136*, 982-983.

5. Franconeri, S.L., and Simons, D.J. (2003). Moving and looming stimuli capture attention. Percept Psychophys *65*, 999-1010.

6. Koch, C., and Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. Trends Cogn Sci *11*, 16-22.

7. Kanai, R., Tsuchiya, N., and Verstraten, F.A. (2006). The scope and limits of top-down attention in unconscious visual processing. Curr Biol *16*, 2332-2336.

8. Yantis, S., and Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. J Exp Psychol Hum Percept Perform *10*, 601-621.

9. Yantis, S., and Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. J Exp Psychol Hum Percept Perform *16*, 121-134.

10. Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res *40*, 1489-1506.

11. Itti, L., and Koch, C. (2001). Computational modelling of visual attention. Nat Rev Neurosci *2*, 194-203.

12. Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. Vision Res *39*, 3157-3163.

13. Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. Nature *290*, 91-97.

14. Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol *4*, 219-227.

15. Treisman, A.M., and Gelade, G. (1980). A feature-integration theory of attention. Cognit Psychol *12*, 97-136.

16. Wolfe, J.M., Cave, K.R., and Franzel, S.L. (1989). Guided search: an alternative to the feature integration model for visual search. J Exp Psychol Hum Percept Perform *15*, 419-433.

17. Duncan, J., and Humphreys, G.W. (1989). Visual search and stimulus similarity. Psychol Rev *96*, 433-458.

18. Nakayama, K., and Mackeben, M. (1989). Sustained and transient components of focal visual attention. Vision Res *29*, 1631-1647.

19. Posner, M.I. (1980). Orienting of attention. Q J Exp Psychol *32*, 3-25.

20. Esteves, F., Parra, C., Dimberg, U., and Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. Psychophysiology *31*, 375-385.

21. Ohman, A., and Soares, J.J. (1998). Emotional conditioning to masked stimuli: expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. J Exp Psychol Gen *127*, 69-82.

22. Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? Curr Opin Neurobiol *15*, 188-196.

23. Regan, D., and Vincent, A. (1995). Visual processing of looming and time to contact throughout the visual field. Vision Res *35*, 1845-1857.

24. Jonides, J., and Irwin, D.E. (1981). Capturing attention. Cognition *10*, 145-150.

25. Bannerman, R.L., Milders, M., de Gelder, B., and Sahraie, A. (2009). Orienting to threat: faster localization of fearful facial expressions and body postures revealed by saccadic eye movements. Proc Biol Sci *276*, 1635-1641.

26. Kentridge, R.W., Nijboer, T.C., and Heywood, C.A. (2008). Attended but unseen: visual attention is not sufficient for visual awareness. Neuropsychologia *46*, 864-869.

27. Georgeson, M. (1997). Vision and action: you ain't seen nothin' yet. Perception *26*, 1-6.

28. Milner, A.D., and Goodale, M.A. (1993). Visual pathways to perception and action. Prog Brain Res *95*, 317-337.

29. Milner, A.D., and Goodale, M.A. (2006). The visual brain in action, 2nd Edition, (Oxford ; New York: Oxford University Press).

30. Aglioti, S., DeSouza, J.F., and Goodale, M.A. (1995). Size-contrast illusions deceive the eye but not the hand. Curr Biol *5*, 679-685.

31. Bridgeman, B. (1981). Cognitive factors in subjective stabilization of the visual world. Acta Psychol (Amst) *48*, 111-121.

32. Daprati, E., and Gentilucci, M. (1997). Grasping an illusion. Neuropsychologia *35*, 1577-1582.

33. Haffenden, A.M., and Goodale, M.A. (1998). The effect of pictorial illusion on prehension and perception. J Cogn Neurosci *10*, 122-136.
34. Newcombe, F., Ratcliff, G., and Damasio, H. (1987). Dissociable visual and spatial impairments following right posterior cerebral lesions: clinical, neuropsychological and anatomical evidence. Neuropsychologia *25*, 149-161.

35. Brainard, D.H. (1997). The Psychophysics Toolbox. Spat Vis *10*, 433-436.

36. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis *10*, 437-442.

37. Zhaoping, L. (2008). Attention capture by eye of origin singletons even without awareness--a hallmark of a bottom-up saliency map in the primary visual cortex. J Vis *8*, 1 1-18.

38. Jiang, Y., Costello, P., Fang, F., Huang, M., and He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. Proc Natl Acad Sci U S A *103*, 17048-17052.

39. Bartolomeo, P., Decaix, C., and Sieroff, E. (2007). The phenomenology of endogenous orienting. Conscious Cogn *16*, 144-161.

APPENDIX A

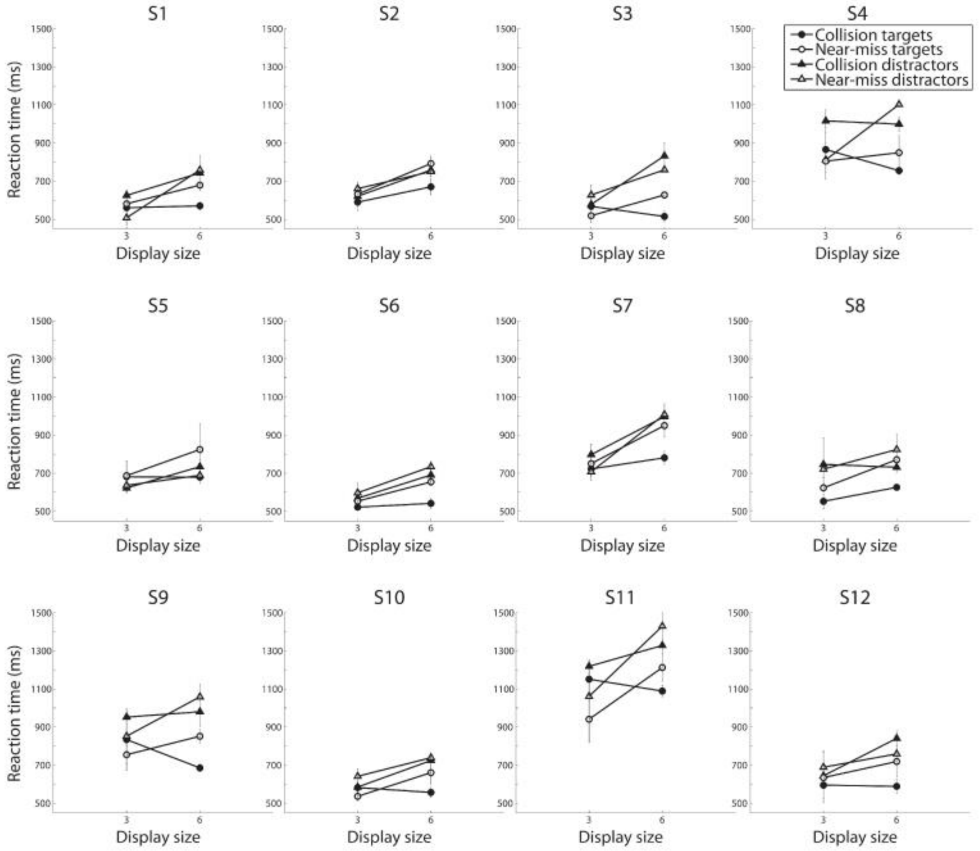

Figure S1 – Mean correct response times and error rates for individual subjects in Experiment 1.

Collision targets and near-miss targets represent trials where the final position of the looming items coincided with the location of the target ovals. Collision distractors and near-miss distractors represent trials where the final position of the looming item was at a location away from the location of the target oval. Increased search efficiency was evident in the rate of search as indexed by search slopes: mean rates of search were fastest for collision targets (-1.7 ms/item), and were much slower for near-miss targets (46.1 ms/item), collision distractors (42.4 ms/item) and near-miss distractors (57.8 ms/item). A significant three-way interaction of position, path, and display size (F(2,7) = 7.56, p = 0.019, MSE = 2005.71) indicates significant differences in the slopes across the four path/position combinations. Error bars represent s.e.m.

APPENDIX B

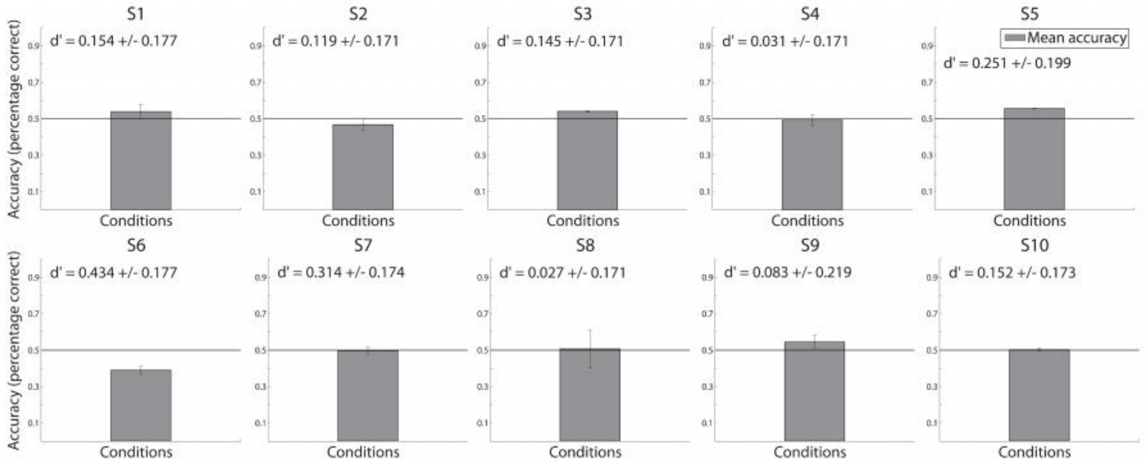

Figure S2 – Mean accuracies for discriminating the trajectory of the looming stimuli for individual subjects from Experiment 2A.

Participants were instructed to fixate at the center, attend to the looming items and identify whether the exhibited looming trajectory was closer to a collision or near-miss with their head.

APPENDIX C

A video clip of a typical trial is available online at:
http://depts.washington.edu/viscog/Research/S3.mov

Enhanced Memory for Scenes Presented at Behaviorally Relevant Points in Time

Jeffrey Y. Lin*, Amanda D. Pype, Scott O. Murray, Geoffrey M. Boynton

Department of Psychology, University of Washington, Seattle, Washington, United States of America

Abstract

The ability to remember a briefly presented scene depends on a number of factors, such as its saliency, novelty, degree of threat, or behavioral relevance to a task. Here, however, we show that the encoding of a scene into memory may depend not only on *what* the scene contains but also *when* it occurs. Participants performed an attentionally demanding target detection task at fixation while also viewing a rapid sequence of full-field photographs of urban and natural scenes. Participants were then tested on whether they recognized a specific scene from the previous sequence. We found that scenes were recognized reliably only when presented concurrently with a target at fixation. This is evidence of a mechanism where traces of a visual scene are automatically encoded into memory at behaviorally relevant points in time regardless of the spatial focus of attention.

Citation: Lin JY, Pype AD, Murray SO, Boynton GM (2010) Enhanced Memory for Scenes Presented at Behaviorally Relevant Points in Time. PLoS Biol 8(3): e1000337. doi:10.1371/journal.pbio.1000337

Academic Editor: Manfred Fahle, Bremen University, Germany

Received August 4, 2009; **Accepted** February 11, 2010; **Published** March 16, 2010

Funding: This work was funded by National Institutes of Health (NIH) RO1 grant EY12925 to GMB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: RSVP, rapid serial visual presentation

* E-mail: jytlin@u.washington.edu

Introduction

Photographs of urban and natural scenes can be perceived and understood very quickly. However, to form a memory of a scene requires substantially more processing time. The dissociation between scene perception and memory has been shown by using rapid serial visual presentation (RSVP) to present a series of images and measuring detection and recognition performance while manipulating exposure duration [1]. These studies have shown that scene understanding requires 100 ms or less while memory formation requires at least an additional 300 ms of processing [1–5]. The amount of time required for memory formation is dependent on a number of factors; for example, it may take longer to form a memory if images from the set being remembered are highly confusable and similar [4].

There are a number of factors that can increase the memorability of a scene. For example, any feature that increases its "distinctiveness" or novelty—from low-level image features (e.g., a low contrast foggy scene among high contrast daylight scenes) to high-level semantic information—can lead to enhanced scene memory [6,7]. Novelty is often believed to transiently increase attention, which leads to enhanced memory—a contention supported by experiments suggesting that spatial attention is necessary for a visual item to be encoded into memory [8–11]. In addition, the processing of novel events is known to rely on unique neural processing [12–17].

Although particularly salient or distinctive information in a scene enhances scene memory, we hypothesized that scene memory would also be enhanced at specific moments in time. A clear example is "flashbulb memory," where details of the context in which people experience shocking news are stored into long-term memory such as where they were, what they were doing, and with whom they were [18]. This suggests the hypothesis that there may be a mechanism in which unattended (but not necessarily physically salient, novel, or threatening) information is implicitly encoded at behaviorally relevant points in time. We explored this hypothesis by testing participants' ability to recognize a particular scene as a member of a sequence of rapidly presented scenes while performing a demanding detection task at fixation. We found that recognition memory was enhanced for test scenes presented concurrently with an unrelated target at fixation. This is evidence of a mechanism where traces of a visual scene are automatically encoded into memory at behaviorally relevant points in time—operationally defined as a point of time that is important for the future execution or completion of an auditory or visual task—regardless of the spatial focus of attention.

Results

Experiment 1

We adapted a standard RSVP task [19] into an RSVP recognition memory task similar to other paradigms used to measure recognition memory for scenes [1,20–22]. In Experiment 1, after being familiarized with a large set of photographs of natural and urban scenes, participants viewed a sequence of 16 scenes presented in an RSVP. Each sequence was then followed by a single test scene in which participants were asked whether they recognized the test scene from the previous RSVP sequence. A typical display sequence is shown in Figure 1. Baseline or chance performance on this task was 50%. Results for the scene recognition task are shown in the grey bar in Figure 2. A *t* test showed that participants performed no better than chance,

Author Summary

What determines whether a visual scene is remembered or forgotten? The ability to remember a briefly presented scene depends on a number of factors, such as its saliency, novelty, degree of threat, or relevance to a behavioral outcome. Generally, attention is thought to be key, in that you can only remember the part of a visual scene you were paying attention to at any given moment. Here, we show that memory for visual scenes may not depend on your attention or *what* a scene contains, but *when* the scene is presented. In this study, attention to one task enhances recognition performance for scenes in a second task only in situations when the first task has behavioral relevance. Our results suggest a mechanism where traces of a visual scene are automatically encoded into memory, even though the scene is not the spatial focus of attention.

51.32%±4.03%, $t(11) = 0.3079$, $p = 0.7639$. Here, participants were unable to recognize whether or not a specific test scene had just appeared in the prior sequence, suggesting a previously unknown difficulty in recognizing a familiar and meaningful scene from short-term memory.

Experiment 2

In Experiment 2, the same set of scenes was presented, but attention was directed to a demanding task at fixation where the goal was to identify a white target letter among a stream of black distractor letters. As in Experiment 1, one scene was presented immediately after each sequence for the recognition test. Mean performance on the letter identification task in Experiment 2 was 95.22%±1.09%, suggesting that participants were complying with instructions to focus their attentional resources on the fixation task. Results for the scene recognition task in Experiment 2 are shown in the white and black bars in Figure 2. The black bar shows recognition performance for scenes presented during distractor frames (black letters). For scenes presented behind black, non-target letters, performance remained at chance—52.49%±1.66%, $t = 0.5951$, $p = 0.5638$.

Surprisingly, scene recognition was significantly greater than chance for test scenes presented concurrently with white target letters (white bar in Figure 2, 67.21%±3.82%). A paired-samples t test reveals a significant difference between recognition task accuracy for test scenes that had previously been presented with black distractor letters versus white target letters, suggesting that scenes presented concurrently with white target letters were remembered better, $t(10) = 2.746$, $p = 0.021$. An additional remarkable feature of Experiment 2 was that participants claimed to have no awareness of their enhanced performance. In debriefing after Experiment 2, all participants claimed that they could not perform the scene recognition task despite performing near 70% on target-present test scenes.

Experiment 3

We next explored whether this improved performance for scene recognition at the time of target detection was specific to detecting visual targets. Participants performed an auditory target detection task while viewing sequences of scenes as in Experiments 1 and 2. Displays and timing parameters were identical to Experiment 2 except that the alphabetical letters were removed from the scenes and replaced with a fixation marker. With every scene, a baseline auditory tone was presented and a unique tone was designated as the target. Mean performance on the auditory detection task was 90.15%±8.19%, which suggests that participants were complying with instructions to focus their attentional resources on the auditory task.

Scene recognition accuracy for Experiment 3 is presented in Figure 3. Similar to Experiment 2, participants performed near chance levels for scenes presented concurrently with distractor tones, 53.59%±1.65%, $t(10) = 0.7290$, $p = 0.4827$. However, performance

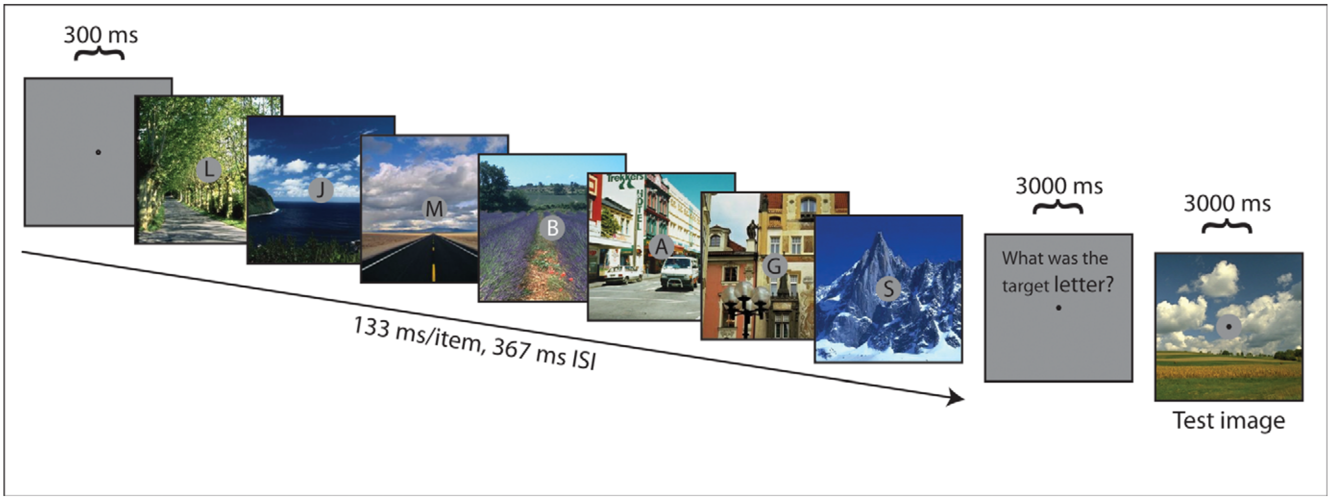

Figure 1. Typical display sequence. Trials were initiated by participants by pressing any letter keys on the keyboard. In Experiment 1, letters at fixation were removed from the displays and participants were instructed to memorize a series of 16 scenes, then to recall whether they recognized a test scene from the RSVP stream (133 ms/item, 367 ms ISI). In Experiment 2, participants were instructed to type the letter key corresponding to the identity of the white target letter for the current trial, then were presented with a test scene and had to recall whether they recognized the scene from the RSVP stream. In Experiment 3, the RSVP letter task was replaced with an auditory task. In Experiment 4, participants received the exact same displays as Experiment 2; however, they were instructed to ignore the letters at fixation and only perform the scene recognition memory task. Prior to testing in every experiment, participants performed a practice block of 24 trials. Each participant was tested for a total of 240 trials, in 10 blocks of 24 trials. Blocks were separated by brief breaks.

doi:10.1371/journal.pbio.1000337.g001

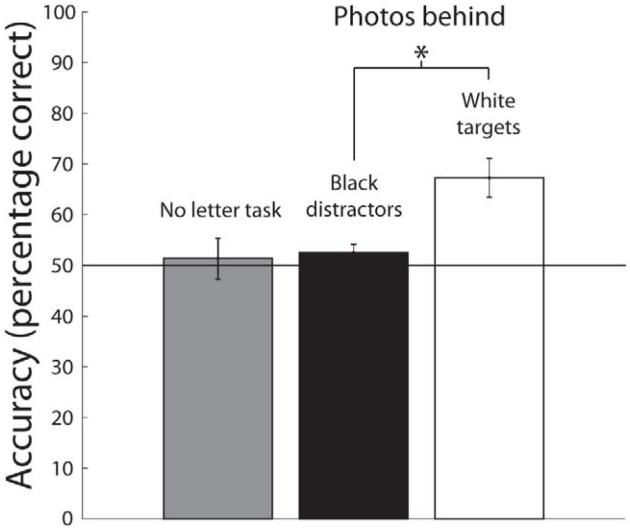

Figure 2. Mean recognition task accuracy for Experiments 1 and 2. Gray bar: results from Experiment 1 where there was no central fixation task. Recognition accuracy was not significantly different from chance performance, 51.32%±4.03%, $t(11) = 0.3079$, $p = 0.7639$. Black bar: results for test scenes presented behind black distractor letters in the target detection task in Experiment 2. Recognition accuracy was again not significantly different from chance performance, 52.49%±1.66%, $t = 0.5951$, $p = 0.5638$. White bar: results for scenes presented concurrently with white target letters in the target detection task in Experiment 2. Recognition accuracy was 67.21%±3.82%. A paired-samples *t* test revealed a significant difference between recognition task accuracy for test scenes that had been previously presented with black distractor letters versus white target letters, suggesting that scenes presented concurrently with white target letters were remembered better, $t(10) = 2.746$, $p = 0.021$. Error bars represent s.e.m. (*$p<0.05$).
doi:10.1371/journal.pbio.1000337.g002

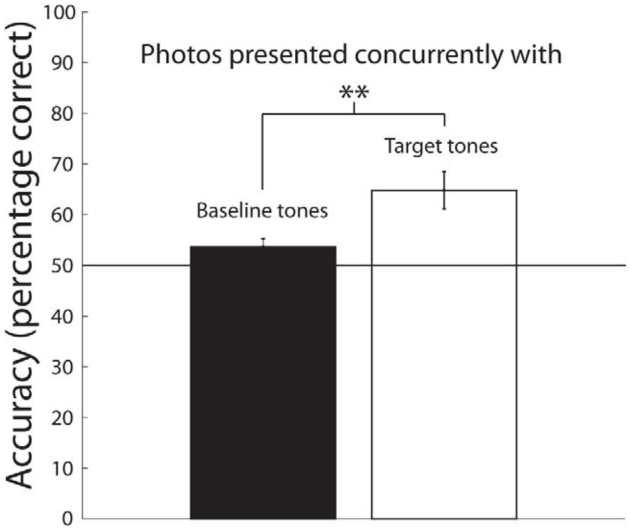

Figure 3. Mean recognition task accuracy for Experiment 3. Photos presented with baseline tones represent trials where the tested scene matched a scene that was presented concurrently with baselines tones in the RSVP stream. Conversely, photos presented concurrently with target tones represent trials where the tested scene matched the scene that was presented with the unique tone in the RSVP stream. In the dual-task condition, recognition accuracy for photos presented concurrently with baseline tones was 53.59%±1.65%, while recognition accuracy for photos presented concurrently with target tones was 64.78%±3.69%. A *t* test reveals that recognition accuracy for photos presented concurrently with baseline tones was not significantly higher than chance levels, $t(10) = 0.7290$, $p = 0.4827$. A paired-samples *t* test revealed a significant difference between recognition task accuracy for test scenes that had been presented concurrently with baseline tones versus target tones, suggesting that scenes presented concurrently with target tones were better encoded into memory, $t(10) = 3.573$, $p = 0.005$. Error bars represent s.e.m. (**$p<0.01$).
doi:10.1371/journal.pbio.1000337.g003

for scenes presented concurrently with target tones were more accurately encoded into memory, 64.78%±3.69%, $t(10) = 3.573$, $p = 0.005$. This shows that enhanced scene encoding occurs for targets detected across modalities, suggesting that the concept of "behavioral relevance in time" is independent of modality.

Experiment 4

In both Experiments 2 and 3, the attended targets were perceptually novel compared to distractor stimuli. Thus, enhanced encoding of scenes during target presentation may be simply due to the physical novelty of the stimuli and not due to performing the detection task. To test this, we used stimuli identical to Experiment 2 including the letter stream at fixation, but participants were instructed to ignore the letters and only perform the scene recognition memory task.

Given that the white letter serves as a perceptually novel event, one might expect enhanced performance for scenes presented concurrently with the novel event. However, recognition performance (shown in Figure 4) was at chance for both test scenes presented concurrently with black distractor letters and with novel white letters, $t(14) = 0.6798$, $p = 0.5077$, and $t(14) = 0.8373$, $p = 0.4165$, respectively. A paired-samples *t* test revealed no significant differences for test scenes presented concurrently with black letters (52.89%±1.33%) and novel white letters (53.13%±3.96%), $t(14) = 0.1494$, $p = 0.8834$, suggesting that the enhanced performance in prior experiments was not simply due to the perceptual novelty of the physical stimulus.

Together, these four experiments demonstrate that at behaviorally relevant points in time—operationally defined as a point of time that is important for the future execution or completion of an auditory or visual task—a memory trace of the visual field is automatically encoded into memory, enhancing later recognition of information even at unattended regions of visual space. This "screen capture" mechanism is likely to play an important role in the retrospective analysis of important events.

Discussion

A defining characteristic of the human visual system is its ability to rapidly extract details of a scene, but it takes substantially longer to encode a scene into memory [1,4]. However, recognition memory for scenes is remarkably good when given sufficient encoding time [23]. Traditionally, the encoding of pictures into memory has been studied using single-task, undivided attention paradigms exploring the effects of stimulus duration and visual and conceptual masking on effective encoding and later memory. Consequently, less is known about memory encoding under conditions of reduced attention.

What determines whether an item is remembered or forgotten? It has been shown that observers are very poor in discriminating or recognizing obvious and significant changes in scenes unless they happen to be attending to the item that was changed [24,25]. As an extension of this, it is generally believed that focused spatial attention is necessary for a visual item to be encoded into memory

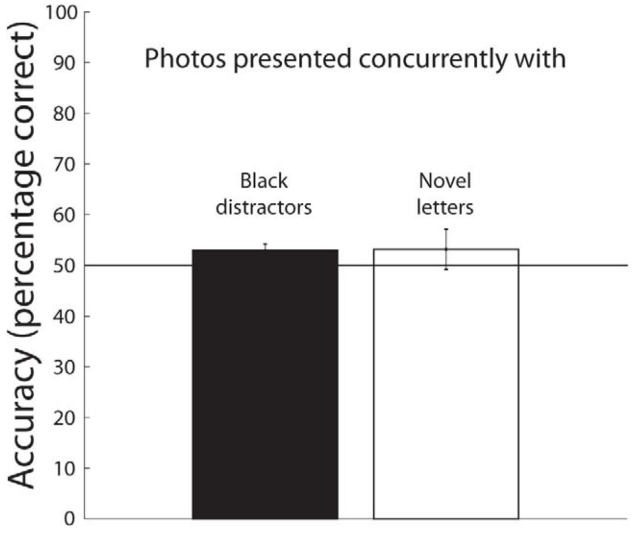

Figure 4. Mean recognition task accuracy for Experiment 4. Photos behind black distractors represent trials where the tested scene matched one of the scenes presented concurrently with a black letter during the RSVP stream. Conversely, photos behind white targets represent trials where the tested scene matched the scene presented concurrently with the white target letter during the RSVP. Displays were identical to Experiment 2; however, participants were instructed to direct their attention to the letters at fixation but only perform the recognition memory task. Given that the white letter serves as a perceptually novel event, one might expect enhanced performance for scenes presented concurrently with the novel event. However, recognition performance was at chance for both test scenes presented concurrently with black distractor letters and with novel white letters, $t(14) = 0.6798$, $p = 0.5077$, and $t(14) = 0.8373$, $p = 0.4165$, respectively. A paired-samples t test revealed no significant differences for test scenes presented concurrently with black letters (52.89%±1.33%) and novel white letters (53.13%±3.96%), $t(14) = 0.1494$, $p = 0.8834$, suggesting that the enhanced performance in prior experiments was not simply due to the perceptual novelty of the physical stimulus.
doi:10.1371/journal.pbio.1000337.g004

[8–11]. In the present studies, however, even focal attention on the scenes in Experiment 1 was not sufficient to maintain familiar scenes in short-term memory.

In sharp contrast, in Experiment 2, when spatial attention was directed towards fixation on an attentionally demanding task, the presentation of a target item resulted in enhanced recognition memory for the scene presented concurrently with the target in the background. This result suggests a new mechanism that may play a role in determining what and when information about a scene is encoded into memory. A counterintuitive feature of this enhanced recognition memory effect is that it occurs in spite of the known effects of focusing of spatial attention around a target item [26,27]. Our results indicate that target detection, engagement, or processing has a strong, non-stimulus-specific influence on memory formation—the enhanced encoding into memory of all items that are temporally coincident with a behaviorally relevant target event. The data suggest that behaviorally relevant points in time trigger a "temporal novelty" effect on memory encoding that appears to be a sufficient prerequisite for the successful encoding of visual stimuli into memory under conditions of reduced attention [28–30].

It is unlikely that this non-stimulus-specific influence on memory formation was due to the attentional blink [19,31] suppressing scenes presented after focal targets were identified; indeed, recognition memory for scenes presented immediately before or after the temporal positions of the targets was still at chance. Moreover, the rate of presentation (two pictures/s) is considerably slower than rates that produce an attentional blink. In addition, recognition memory for the scenes presented before or after the temporal positions of the targets being at chance also suggests that the effects were not due to a general arousal [32] triggered by the onset of a perceptually novel stimulus and thereby increasing recognition memory for all subsequent scenes presented after the targets.

Perceptual learning for task-irrelevant peripheral stimuli can occur when attention is focused away from the peripheral stimuli and towards fixation and these learning effects are greatest for peripheral stimuli presented at the time of foveal target detection [33–35]. These results were surprising because it had generally been assumed that perceptual learning requires attention be focused on the target stimulus being learned. However, even in the absence of attention, it must be necessary for the target stimulus being learned to be encoded into memory for learning to occur. Here, we show that short-term memory for a peripheral scene is enhanced when it is presented at a behaviorally relevant point in time. It seems likely that a version of this "task-related screen capture" is one of the mechanisms that could support the phenomenon of perceptual learning in the absence of attention.

Recently, researchers have shown that repeated presentation of movie clips produces detectable "memory traces" in subsequent resting state activity in cat visual cortex [36]. It is plausible that given a behaviorally relevant point in time, a strong reverberation or memory trace was triggered and the residual of this imprint was being tapped into when performing the scene recognition task.

Finally, one might assume these results suggest that the processes associated with enhanced vividness, memory, and attention for novel events act globally throughout the visual field; however, Experiment 4 suggests that at first glance, perceptual novelty is not the source of these effects. When passively viewing the same displays as Experiment 2 and asked to perform the recognition memory task while ignoring the black distractor letters and novel white target letters, no significant differences were found in recognition performance. Overall, our results suggest a mechanism where traces of a visual scene are automatically encoded into memory at behaviorally relevant points in time regardless of the spatial focus of attention.

Materials and Methods

All participants reported normal or corrected-to-normal visual acuity and gave informed consent to participate in this experiment, which was approved by the University of Washington Human Subjects Institutional Review Board. In every experiment prior to testing, participants performed a practice block of 24 trials. Each participant was then tested for a total of 240 trials, in 10 blocks of 24 trials. Blocks were separated by brief breaks.

Different participants participated in each of the five experiments. All received financial compensation in one 1 h session. Experiment 1 consisted of 12 participants (10 females, 2 males). Experiment 2 consisted of 11 participants (7 females, 4 males). Experiment 3 consisted of 11 participants (6 females, 5 males). Experiment 4 consisted of 15 participants (11 females, 4 males).

Apparatus and Stimuli

Displays were presented on a 45 cm ViewSonic Graphics Series G90fB monitor at 1024×768 resolution, refreshed at 60 Hz. Participants sat with their eyes approximately 50 cm from the screen. The backgrounds of all displays were gray (15 cd/m^2). Display items consisted of 192 700×700 pixel (28.07 degrees of visual angle) photographs depicting natural or urban scenes from eight distinct categories (i.e., mountains, cityscapes, etc). Scenes were obtained from the LabelMe Natural and Urban Scenes database [37] at 250×250 pixels of resolution, then up-sampled to 700×700 pixels of resolution.

Display items during the experiment were sampled from the 192 scenes with replacement. In each sequence, observers were shown 16 of these scenes at 133 ms per scene, followed by a blank ISI of 367 ms for a SOA of 500 ms.

Scene Recognition Task

All experiments (1, 2, 3, and 4) used the scene recognition task. Following each rapid sequence of 16 full-field scenes, observers were presented with a test scene and asked to recall whether the test scene appeared in the previous RSVP sequence of scenes. The test scene was presented for 3,000 ms or until participants responded to whether they recognized the test scene from the RSVP stream with a "Y" or "N" on the keyboard. In 50% of the trials, the test scene was randomly drawn from the scenes presented in serial positions 9 to 16 of the RSVP; in the other 50% of trials, the test scene was drawn from the set of scenes not shown in the current RSVP stream. When test scenes were drawn from serial positions 9 to 16, there was a random 1/8 chance that the test scene matched the scene presented behind the white target letter in the RSVP stream, meaning that the white target letter task was irrelevant to the secondary recognition memory task and did not predict the test scene participants would be tested on. All scenes were sampled from our database with replacement. Distractors and target letters were embedded in randomly selected scenes over the entire session.

It is important to note that although our scene recognition task is similar to earlier studies that tested picture memory for novel scenes [1,20], our task requires the participant to remember whether an already-familiar test picture appeared in the most recent sequence. Previous studies have used unfamiliar pictures on each trial. We presume that observers would have no difficulty detecting the presence or absence of a familiar scene in a sequence if they knew beforehand what scene to detect [3,4]. In addition to the main result, the last scene in the RSVP sequence was often recognized with higher accuracy, in line with well-known recency effects of memory [38–43] and the fact that the last scene was not conceptually masked by a subsequent item. In Experiment 1, we only tested the second half of scenes presented in the RSVP to maintain consistency with subsequent experiments and therefore do not have data on potential performance differences for the first scene presented in the RSVP sequence. This new recognition memory task that measured participants' ability to encode a familiar set of scenes into short-term memory using RSVP sequences served as a starting point for examining potential temporally related enhancements to the encoding of briefly presented scenes into memory.

Letter Target Identification Task

For the letter detection task (Experiment 2), a gray aperture (1 degree of visual angle) was embedded in the center of each scene and a random alphabetical letter (20 font size) was centered within the aperture. New random letters were embedded into the gray apertures of every scene, with the only requirement being that no duplicate letters could be presented within the same trial. Alphabetical letters were either black (indicating its identity as a distractor) or white (indicating its identity as a target; see Figure 1). In every trial, random black alphabetical letters representing distractors were embedded at central fixation in 15 of the scenes and a random white alphabetical letter representing the target was embedded in 1 scene. White target letters could only appear concurrently with scenes presented in serial positions 9 to 16 to avoid having white target letters presented at the onset of a RSVP stream. Participants were instructed to fixate on a point in the center of the screen and search for and identify a white target letter while memorizing the series of 16 scenes presented in RSVP.

In Experiment 2, immediately following the RSVP, participants were instructed to type the letter key corresponding to the identity of the white target letter for the current trial. Following the response to the letter detection task, participants performed the scene detection task. Participants were instructed to ignore the letter stream in Experiment 4.

Auditory Target Identification Task

The auditory target detection task in Experiment 3 was similar to the letter detection task in Experiment 2 except the alphabetical letters were removed from the apertures centered in the scenes. Instead, an auditory tone was presented with each scene. Tones were sampled at 44,000 Hz, with durations of 50 ms. Baseline tones were presented at 261.50 Hz, while target tones were either 130.75 Hz or 523.0 Hz. Immediately following the RSVP stream, participants were instructed to discriminate the pitch of the unique tone as either lower or higher via key press, then were again presented with a test scene and asked to recall whether they recognized the scene from the RSVP stream.

Acknowledgments

We thank Mary Potter, Ione Fine, and Sung Jun Joo for helpful discussion.

Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: JYTL. Performed the experiments: JYTL ADP. Analyzed the data: JYTL. Wrote the paper: JYTL ADP SOM GMB.

References

1. Potter MC, Levy EI (1969) Recognition memory for a rapid sequence of pictures. J Exp Psychol 81: 10–15.
2. Greene MR, Oliva A (2009) The briefest of glances: the time course of natural scene understanding. Psychol Sci 20: 464–472.
3. Potter MC (1975) Meaning in visual search. Science 187: 965–966.
4. Potter MC (1976) Short-term conceptual memory for pictures. J Exp Psychol [Hum Learn] 2: 509–522.
5. Potter MC (1993) Very short-term conceptual memory. Mem Cognit 21: 156–161.
6. Hunt RR, Worthen J, eds. Distinctiveness and memory. New York: Oxford University Press.
7. Lin PH, Luck SJ (2009) The influence of similarity on visual working memory representations. Vis Cogn 17: 356–372.
8. Awh E, Jonides J (2001) Overlapping mechanisms of attention and spatial working memory. Trends Cogn Sci 5: 119–126.
9. Smyth MM, Scholey KA (1994) Interference in immediate spatial memory. Mem Cognit 22: 1–13.
10. Turk-Browne NB, Yi DJ, Chun MM (2006) Linking implicit and explicit memory: common encoding factors and shared representations. Neuron 49: 917–927.
11. Wolfe JM, Horowitz TS, Michod KO (2007) Is visual attention required for robust picture memory? Vision Res 47: 955–964.
12. Fabiani M, Donchin E (1995) Encoding processes and memory organization: a model of the von Restorff effect. J Exp Psychol Learn Mem Cogn 21: 224–240.
13. Fabiani M, Karis D, Donchin E (1986) P300 and recall in an incidental memory paradigm. Psychophysiology 23: 298–308.
14. Fernandez G, Tendolkar I (2001) Integrated brain activity in medial temporal and prefrontal areas predicts subsequent memory performance: human declarative memory formation at the system level. Brain Res Bull 55: 1–9.

15. Hasselmo ME (1999) Neuromodulation: acetylcholine and memory consolidation. Trends Cogn Sci 3: 351–359.
16. Kirchhoff BA, Wagner AD, Maril A, Stern CE (2000) Prefrontal-temporal circuitry for episodic encoding and subsequent memory. J Neurosci 20: 6173–6180.
17. Parker A, Wilding E, Akerman C (1998) The Von Restorff effect in visual object recognition memory in humans and monkeys. The role of frontal/perirhinal interaction. J Cogn Neurosci 10: 691–703.
18. Brown R, Kulik J (1977) Flashbulb memories. Cognition 5: 73–99.
19. Raymond JE, Shapiro KL, Arnell KM (1992) Temporary suppression of visual processing in an RSVP task: an attentional blink? J Exp Psychol Hum Percept Perform 18: 849–860.
20. Intraub H (1980) Presentation rate and the representation of briefly glimpsed pictures in memory. J Exp Psychol Hum Learn 6: 1–12.
21. Potter MC, Staub A, O'Connor DH (2004) Pictorial and conceptual representation of glimpsed pictures. J Exp Psychol Hum Percept Perform 30: 478–489.
22. Potter MC, Staub A, Rado J, O'Connor DH (2002) Recognition memory for briefly presented pictures: the time course of rapid forgetting. J Exp Psychol Hum Percept Perform 28: 1163–1175.
23. Shepard RN (1967) Recognition memory for words, sentences and pictures. J Verb Learn Behav 6: 156–163.
24. O'Regan JK, Rensink RA, Clark JJ (1999) Change-blindness as a result of 'mudsplashes'. Nature 398: 34.
25. Pashler H (1988) Familiarity and visual change detection. Percept Psychophys 44: 369–378.
26. Handy TC, Mangun GR (2000) Attention and spatial selection: electrophysiological evidence for modulation by perceptual load. Percept Psychophys 62: 175–186.
27. Johnson DN, McGrath A, McNeil C (2002) Cuing interacts with perceptual load in visual search. Psychol Sci 13: 284–287.
28. Desimone R (1996) Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci U S A 93: 13494–13499.
29. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18: 193–222.
30. Huang L, Pashler H (2007) Working memory and the guidance of visual attention: consonance-driven orienting. Psychon Bull Rev 14: 148–153.
31. Chun MM, Potter MC (1995) A two-stage model for multiple target detection in rapid serial visual presentation. J Exp Psychol Hum Percept Perform 21: 109–127.
32. Solomon JA (2004) The effect of spatial cues on visual sensitivity. Vision Res 44: 1209–1216.
33. Seitz AR, Watanabe T (2003) Psychophysics: is subliminal learning really passive? Nature 422: 36.
34. Seitz AR, Yamagishi N, Werner B, Goda N, Kawato M, et al. (2005) Task-specific disruption of perceptual learning. Proc Natl Acad Sci U S A 102: 14895–14900.
35. Watanabe T, Nanez JE, Sasaki Y (2001) Perceptual learning without perception. Nature 413: 844–848.
36. Yao H, Shi L, Han F, Gao H, Dan Y (2007) Rapid learning in cortical coding of visual scenes. Nat Neurosci 10: 772–778.
37. Oliva A, Torralba A (2001) Modeling the shape of a scene: a holistic representation of the spatial envelope. Int J Comp Vis M 42: 145–175.
38. Bonanni R, Pasqualetti P, Caltagirone C, Carlesimo GA (2007) Primacy and recency effects in immediate free recall of sequences of spatial positions. Percept Mot Skills 105: 483–500.
39. Brooks BM (1999) Primacy and recency in primed free association and associative cued recall. Psychon Bull Rev 6: 479–485.
40. Demaree HA, Shenal BV, Everhart DE, Robinson JL (2004) Primacy and recency effects found using affective word lists. Cogn Behav Neurol 17: 102–108.
41. Labbie S (1973) Primacy-recency effects in impression formation and congruity-incongruity of stimulus material. Percept Mot Skills 37: 275–278.
42. Mondor TA, Morin SR (2004) Primacy, recency, and suffix effects in auditory short-term memory for pure tones: evidence from a probe recognition paradigm. Can J Exp Psychol 58: 206–219.
43. Tan L, Ward G (2000) A recency-based account of the primacy effect in free recall. J Exp Psychol Learn Mem Cogn 26: 1589–1625.

Rapid and reflexive feature-based attention

Jeffrey Y. Lin University of Washington, Seattle, WA, USA
Bjorn Hubert-Wallander University of Washington, Seattle, WA, USA
Scott O. Murray University of Washington, Seattle, WA, USA
Geoffrey M. Boynton University of Washington, Seattle, WA, USA

Performance on a visual task is improved when attention is directed to relevant spatial locations or specific visual features. Spatial attention can be directed either voluntarily (endogenously) or automatically (exogenously). However, feature-based attention has only been shown to operate endogenously. Here, we show that an exogenous cue to a visual feature can lead to improved performance in visual search. Response times were measured as subjects detected or discriminated a target oval among an array of disks, each with a unique color. An uninformative colored cue was flashed at the beginning of each trial that sometimes matched the location and/or color of the target oval. Subjects detected or discriminated the target faster when the color of the cue matched the color of the target, regardless of the cue's location relative to the target. Our results suggest evidence for a previously unknown exogenous cuing mechanism for feature-based attention.

Keywords: attention, spatial vision, search

Citation: Lin, J. Y., Hubert-Wallander, B., Murray, S. O., & Boynton, G. M. (2011). Rapid and reflexive feature-based attention. *Journal of Vision, 11*(12):12, 1–9, http://www.journalofvision.org/content/11/12/12, doi:10.1167/11.12.12.

Introduction

Our everyday visual experience is strongly affected by attention. Visual attention can enhance or prioritize the processing of specific stimuli over the overwhelming number of other sensory inputs by selecting spatial locations (Bisley & Goldberg, 2010; Carrasco, 2011; Chun, Golomb, & Turk-Browne, 2011; Desimone & Duncan, 1995; Egeth & Yantis, 1997; Johnston & Dark, 1986; Kinchla, 1992; Logan, 1996; Pashler, Johnston, & Ruthruff, 2001; Posner, 1980; Posner & Rothbart, 2007), features (Boynton, Ciaramitaro, & Arman, 2006; Maunsell & Treue, 2006; McAdams & Maunsell, 2000; Saenz, Buracas, & Boynton, 2002; Serences & Boynton, 2007; Treue & Martínez Trujillo, 1999), objects (Blaser, Pylyshyn, & Holcombe, 2000), and even time (Lin, Pype, Murray, & Boynton, 2010; Nobre, Allison, & McCarthy, 1998; Swallow & Jiang, 2010). In the past, studies have focused on spatial attention and its derivatives—volitional (endogenous) and reflexive (exogenous) allocation of spatial attention. Feature-based attention, however, has typically been studied using paradigms in which participants voluntarily deploy attention toward a particular visual feature. Attending to a feature such as color or motion in this way produces a global facilitation of processing for stimuli containing that feature throughout the visual field (Liu & Mance, 2011; Sàenz, Buracas, & Boynton, 2003; White & Carrasco, 2011). Given that numerous studies have shown that exogenous cues can enhance sensitivity to visual input at a cued location (Bisley & Goldberg, 2010; Carrasco, 2011; Chun et al., 2011; Desimone & Duncan, 1995; Egeth & Yantis, 1997; Johnston & Dark, 1986; Kinchla, 1992; Logan, 1996; Pashler et al., 2001; Posner, 1980; Posner & Rothbart, 2007), we hypothesized that salient features of an exogenous cue might automatically produce feature-based attentional effects that are independent of location-based cuing effects.

To test this hypothesis, we used a visual search paradigm preceded by different combinations of exogenous spatial and feature cues. We found that the color feature on the cue can rapidly and reflexively produce search benefits at multiple locations throughout the visual field as quickly as 260 ms after the onset of the cue. These feature-based search benefits are independent of spatially based search benefits and the two can combine to produce search benefits beyond either alone.

Methods

Participants

Twenty-one members of the University of Washington community (10 females and 11 males) received financial compensation or course credit for participating in Experiment 1, 16 participated in Experiment 2 (8 females and 8 males), and 18 participated in Experiment 3 (14 females and 4 males). All reported normal or corrected-to-normal

doi: 10.1167/11.12.12 Received April 27, 2011; published October 18, 2011 ISSN 1534-7362 © ARVO

visual acuity. All subjects gave informed consent to participate in this experiment, which was approved by the University of Washington Human Subjects Institutional Review Board.

Apparatus, stimuli, and procedure

Displays were generated in Matlab (Mathworks) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and presented on a 51-cm (diagonal) Samsung Syncmaster 1100DF CRT monitor at 1024 × 768 resolution, refreshed at 75 Hz. Participants used a chin rest and sat with their eyes 50 cm from the screen. Prior to the start of the experimental trials, participants were asked to discriminate between the 6 colors used in the experiments to ensure adequate color vision. The background of the displays was black (0 cd/m^2). Display items consisted of disks (1.8 degrees of visual angle) filled with 1 of 6 solid colors.

Each trial in the experiments consisted of four stages (see Figure 1 for a schematic illustration). (1) The initial preview display duration was jittered between 250 and 500 ms and consisted of a small fixation marker. (2) This was followed by a colored square cue (0.92 degree of visual angle), presented for 59 ms at one of the item locations eventually presented in the search display, then a blank screen for 200 ms. (3) The colored cue was followed by a visual search display with 3 or 6 items placed randomly in eight possible locations positioned evenly on a circular array around fixation with a radius of 5.03 deg. This search display remained in view until participants responded or 2000 ms elapsed.

In 50% of search displays, a target oval was created by narrowing a circle by 5.5% (1.8 deg to 1.7 deg) along the horizontal dimension or vertical dimension. Trials varied by (1) *target presence*: whether the target oval was present or absent, (2) *spatial validity*: whether the cue's location coincided with the target oval's location, (3) *feature validity*: whether the color of the cue coincided with the color of the target oval, and (4) *display size*: the size of the search array (either 3 or 6 items). The location and color of the cue were counterbalanced and randomized in every block for every participant. Importantly, counterbalancing was done in such a way that the cue locations and colors were completely uninformative with respect to the eventual location of the target in the search displays. For set size 3 in Experiment 1, this meant that, overall, the cue's location, color, or combination of location and color

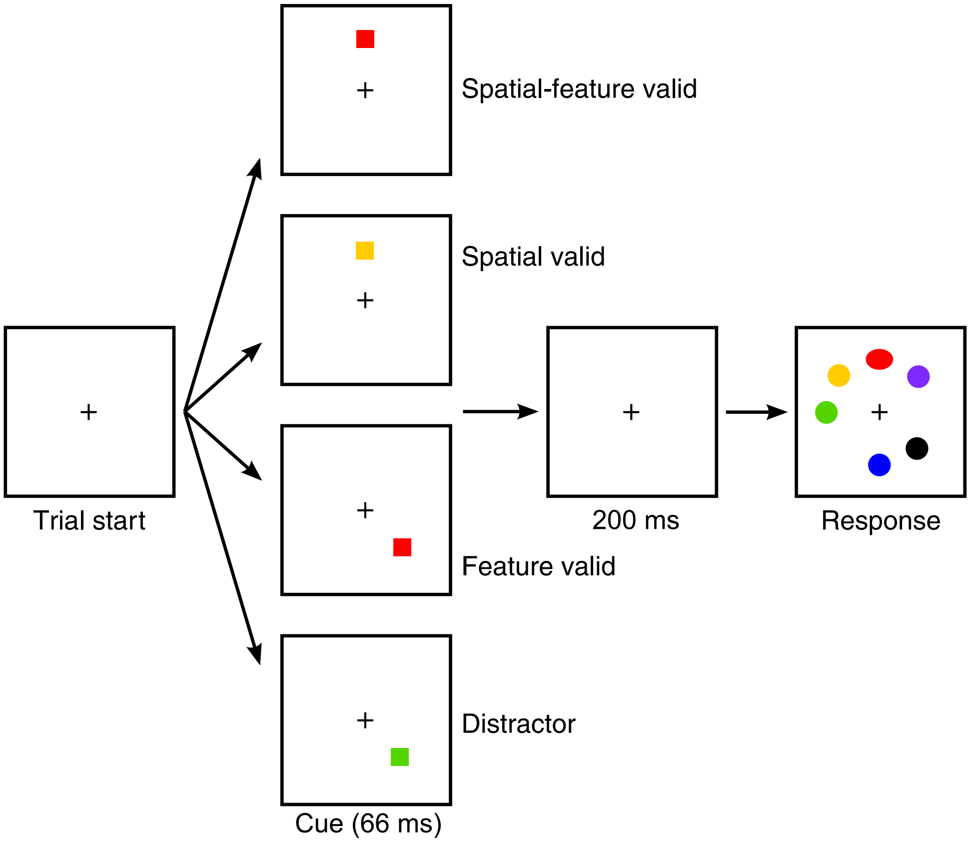

Figure 1. Illustration of a sample visual search sequence in Experiment 1. The displays in each trial consisted of a fixation marker for 250–500 ms, followed by an exogenous cue, then a search display where participants had to locate the presence of an oval among circles. Exogenous cues varied in both spatial and feature validity and were always uninformative relative to the eventual location of the oval target.

coincided with the target oval by chance on 33.3% of the trials. For set size 6 in Experiments 1 and 3, specific breakdowns of trials were generated such that 16.7% of the trials in set size 6 were spatial valid trials and 16.7% of the trials were feature valid trials. Consequently, in set size 6, the cue's location and color both coincided with the target oval by chance on 2.78% of trials.

In Experiments 1 and 2, participants were instructed to search for a target oval (while maintaining fixation at the center fixation point) and to press a key as quickly as possible if the target oval was present. If the oval was not present, participants had to wait until the trial timer elapsed, withholding any response. A small plus sign (correct), minus sign (incorrect), or circle (no response) provided feedback after each trial and was replaced by a fixation marker to serve as the fixation for the following trial and signal the start of the next set of displays. Participants were explicitly informed that target ovals would be present in 50% of the displays and search displays would be preceded by colored cues that were not predictive of the eventual location of the target oval. Participants were instructed to respond as quickly as possible while maintaining an overall accuracy of at least 80%. Each participant was tested for a total of 540 trials, in 9 blocks of 60 trials. Blocks were separated by brief breaks, with a longer break at the halfway point.

In Experiment 2, displays were identical to Experiment 1 except for two critical differences. A set size of 5 was used instead of 3 and 6 in Experiment 1, and the exogenous cues were only presented at fixation, at the location of the fixation marker. Each participant was tested for a total of 1200 trials, in 20 blocks of 60 trials. In both experiments, participants completed 40 practice trials prior to the test trials.

In Experiment 3, displays were identical to Experiment 1; however, the task was changed to a discrimination task and only set size 6 was used. Target ovals were oriented either vertically or horizontally and subjects responded on each trial by pressing "V" or "H" on the keyboard, respectively. Participants were explicitly informed that target ovals would be present in 100% of the displays. Each participant was tested for a total of 1296 trials in 16 blocks of 81 trials. Participants completed 40 practice trials prior to the test trials.

Experiment 1

The aim of Experiment 1 was to measure the effects of uninformative exogenous feature and spatial cues on response times (RTs) when detecting a target across four conditions in a 2 × 2 design (spatial validity × feature validity).

Figure 2A shows mean response times for trials where observers correctly detected target ovals in set size 6. Subjects detected targets fastest in the spatial–feature valid condition (mean = 765 ms, *SD* = 171 ms), followed by the spatial valid (mean = 818 ms, *SD* = 176 ms), feature valid (mean = 876 ms, *SD* = 174 ms), and spatial and feature invalid conditions (mean = 912 ms, *SD* = 174 ms). A repeated-measures 2 (*spatial validity*) × 2 (*feature validity*) ANOVA shows the expected main effect for

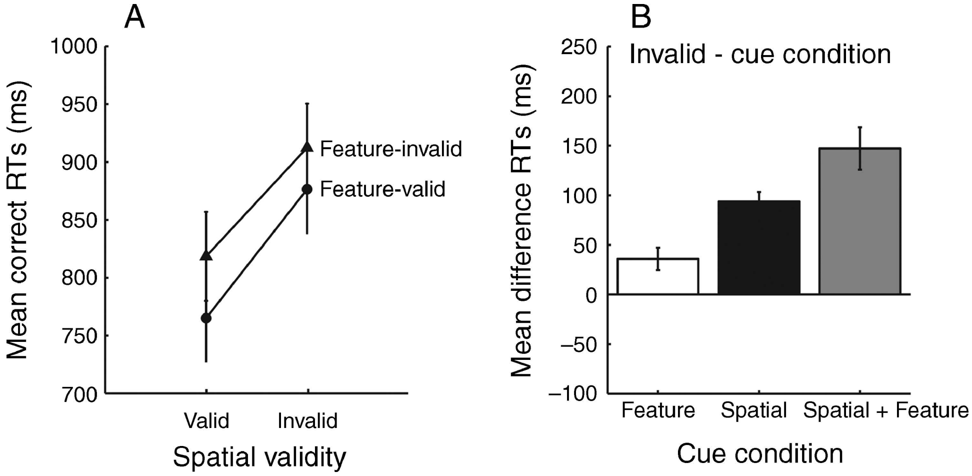

Figure 2. (A) Mean correct response times (RTs) for Experiment 1. Mean correct reaction times suggested that response times to spatial and feature valid cues were faster than response times to spatial and feature invalid cues. Interestingly, there was no significant interaction between spatial and feature validity, indicating that the spatial and feature validity benefits did not interact to produce a multiplicative reduction in reaction times. Error bars represent 1 *SEM*. (B) Mean correct difference reaction times (RTs) for Experiment 1. Cue benefits represent the difference between RTs from both spatially and featurally invalid trials and the three valid cue types (spatially valid, featurally valid, and both spatially and featurally valid). Bonferroni–Holm corrected paired-samples *t*-tests suggested that cues that were both spatially and featurally valid produced significant cue benefits above and beyond cues that were just spatially valid. Error bars represent 1 *SEM*.

spatial validity, $F(1,20) = 56.70$, $p < 0.01$, partial $\eta^2 = 0.74$, but also a significant main effect for feature validity, $F(1,20) = 14.20$, $p < 0.01$, partial $\eta^2 = 0.42$. There was no significant interaction between spatial and feature validity, $F(1,20) = 0.64$, $p = 0.43$, partial $\eta^2 = 0.03$.

Figure 2B shows mean correct difference RTs for Experiment 1. Cue benefits represent the difference (in ms) between RTs from each of the three valid conditions (spatially valid, featurally valid, and both spatially and featurally valid) and the spatially and featurally invalid condition. Five post-hoc paired *t*-tests were conducted to more precisely determine the relationships between the different cue conditions, with the familywise type I error rate held at 0.05 via Bonferroni–Holm *p*-value correction. As expected, each of the valid cue conditions resulted in faster reaction times than the spatial and feature invalid condition (*spatial valid vs. spatial and feature invalid:* $t(20) = 9.78$, corrected $p < 0.001$; *feature valid vs. spatial and feature invalid:* $t(20) = 3.17$, corrected $p < 0.01$; *spatial–feature valid vs. spatial and feature invalid:* $t(20) = 6.88$, corrected $p < 0.001$). Furthermore, while a spatially valid cue produced a larger RT benefit than a featurally valid cue ($t(20) = 4.17$, corrected $p = 0.001$), RTs for spatial–feature valid cues were significantly faster than just spatial valid alone ($t(20) = 2.70$, corrected $p = 0.014$), indicating that the addition of feature validity further facilitated reaction times to targets even when the spatial cue already matched the location of the target. These tests together demonstrate that the main effect of feature validity was driven by feature-based benefits in both spatially valid and spatially invalid trials, not just in one or the other.

Analysis of accuracies indicated no evidence of a speed–accuracy trade-off in set size 6. Average accuracy rates were very high (spatial and feature invalid = 94.55%, *SEM* = 1.18%; spatial valid = 94.64%, *SEM* = 1.18; feature valid = 94.71%, *SEM* = 1.38%; spatial and feature valid = 96.66%, *SEM* = 1.80%). A one-way ANOVA revealed that accuracy did not vary significantly across the four conditions, $F(3,60) = 1.090$, $p = 0.36$, as expected. These results replicate previous findings that search items that coincide with the location of a preceding exogenous cue receive prioritized processing in the visual search that follows (Giordano, McElree, & Carrasco, 2009; Posner, 1980; Posner & Petersen, 1990; Schreij, Owens, & Theeuwes, 2008; Schreij, Theeuwes, & Olivers, 2010; Yantis, 1993; Yantis & Egeth, 1999; Yantis & Jonides, 1984, 1990). However, these results also reveal a novel search benefit when the target and the cue share a feature (color), regardless of the relative location of the two items. This would appear to be the result of the exogenous cue producing a global spread of feature-based attention that operates alongside normal spatial cuing mechanisms. Furthermore, the lack of a significant interaction effect between spatial and feature validity suggests that validity in both cue dimensions results in the reaction time benefit that would be expected by summing the individual benefits of both spatially valid and featurally valid cues. In other words, the spatial- and feature-based cuing benefits appear to originate independently of each other.

In contrast to set size 6, results from set size 3 yielded relatively small and inconclusive effects. Here, targets were detected fastest in the spatial valid condition (mean = 701 ms, *SD* = 132 ms), followed by the spatial–feature valid (mean = 719 ms, *SD* = 126 ms), feature valid (mean = 731 ms, *SD* = 122 ms), and spatial and feature invalid conditions (mean = 758 ms, *SD* = 132 ms). A repeated-measures 2 (*spatial validity*) × 2 (*feature validity*) ANOVA was conducted, showing trends consistent with the set size 6 data but no significance at the $\alpha = 0.05$ criterion. There was no statistically significant main effect of spatial validity, $F(1,20) = 3.90$, $p = 0.06$, partial $\eta^2 = 0.16$, or feature validity, $F(1,20) = 0.10$, $p = 0.75$, partial $\eta^2 = 0.01$, nor was there a significant interaction between the two, $F(1,20) = 1.74$, $p = 0.20$, partial $\eta^2 = 0.08$. In order to discern any differences between the conditions, all possible post-hoc paired *t*-tests were conducted on RTs from the four cue types, with familywise type I error rate held at 0.05 via Bonferroni–Holm *p*-value correction. As might be expected from the lack of any effects in the ANOVA, none of the *t*-tests revealed significant differences between the different cue types (smallest corrected $p = 0.24$). Analysis of accuracies indicated no evidence of a speed–accuracy trade-off in set size 3. Average accuracy rates were very high (spatial and feature invalid = 96.00%, *SEM* = 1.00%; spatial valid = 96.00%, *SEM* = 1.50%; feature valid = 94.40%, *SEM* = 1.70%; spatial and feature valid = 96.00%, *SEM* = 1.70%). A one-way ANOVA revealed that accuracy did not vary significantly across the four conditions, $F(3,60) = 0.311$, $p = 0.82$, as expected.

These results for a set size of 3 might be considered a concern for the hypothesis that exogenous cues in our paradigm produced a rapid and reflexive spread of feature-based attention since we observed no discernable RT benefit in response to cues that matched the target color. However, our choice of a three-item search display might not have allowed any possible benefit from a reflexive spread of feature-based attention to manifest even if it existed. In particular, the aspect ratios of our oval targets may have "popped out" of the search array when present in our set size 3 displays, negating any potential cue benefits. There is some evidence for this hypothesis, as we found no significant differences in mean RTs between *any* of our four cuing conditions. Indeed, we did not achieve statistical significance for the highly studied effect of exogenous spatial cuing (Posner, 1980; Yantis & Jonides, 1984).

Experiment 2

Our results thus far have suggested the existence of an exogenous feature-based cuing benefit that is spatially independent. However, the cues that produced these benefits always occurred at potential target locations (i.e., confined to the ring of possible circle/oval locations). It is possible, therefore, that featurally valid cues occurring at locations where a target never appears may produce minimal or no benefits. To test this, we conducted a second experiment that was identical to the first except that the set size was fixed at 5 objects and the cues always appeared at fixation rather than at one of the object locations in the search display. A paired *t*-test revealed no significant difference in RTs between trials where the cue was featurally valid (mean = 650 ms, *SD* = 76 ms) relative to trials where the cue was featurally invalid (mean = 647 ms, *SD* = 63 ms), $t(14) = 0.32$, $p = 0.74$, indicating that there may be no discernable feature-based cuing effect when cues are presented at fixation. Average accuracy rates were very high and suggested no speed–accuracy trade-offs (feature invalid = 97.11%, *SEM* = 0.99%; feature valid = 95.67%, *SEM* = 0.55%). A paired *t*-test revealed that accuracy did not vary significantly across the two conditions, $t(14) = 1.78$, $p = 0.10$, as expected. These results suggest that while the benefits of a valid exogenous feature cue are spatially independent, the cue itself may have to be in a target-relevant location to achieve this benefit.

Experiment 3

If the effects observed in Experiment 1 are due to the reflexive orienting of feature-based attention to an object exhibiting a matching feature of an exogenous cue, then the effects should generalize across different tasks. The aim of Experiment 3 was to test for effects of reflexive feature-based attention on visual search in a discrimination task where attention must be focused on the target stimulus to complete the task. The stimulus configurations were identical to Experiment 1 except that only a set size of 6 was presented.

Figure 3A shows mean response times (RTs) for trials where observers correctly discriminated oval orientations. Similarly to Experiment 1, subjects chose the correct orientation fastest in the spatial–feature valid condition (mean = 800 ms, *SD* = 87 ms), followed by spatial valid (mean = 813 ms, *SD* = 91 ms), feature valid (mean = 849 ms, *SD* = 110 ms), and spatial and feature invalid conditions (mean = 873 ms, *SD* = 116 ms).

A repeated-measures 2 (*spatial validity*) × 2 (*feature validity*) ANOVA revealed the expected main effect for spatial validity, $F(1,17) = 29.68$, $p < 0.001$, partial $\eta^2 = 0.64$, and also the same significant main effect for feature validity found in Experiment 1, $F(1,17) = 12.04$, $p < 0.003$, partial $\eta^2 = 0.42$. Further replicating the results from Experiment 1, there was no significant interaction between spatial and feature validity, $F(1,17) = 1.23$, $p = 0.28$, partial $\eta^2 = 0.07$.

Figure 3B shows mean correct difference RTs for Experiment 3. Cue benefits represent the difference (in ms) between RTs from each of the three valid conditions (spatially valid, featurally valid, and both spatially and featurally valid) and the spatially and featurally invalid condition. Five post-hoc paired *t*-tests were conducted to more precisely determine the relationships between the different cue conditions, with the familywise type I error rate held at 0.05 via Bonferroni–Holm *p*-value correction.

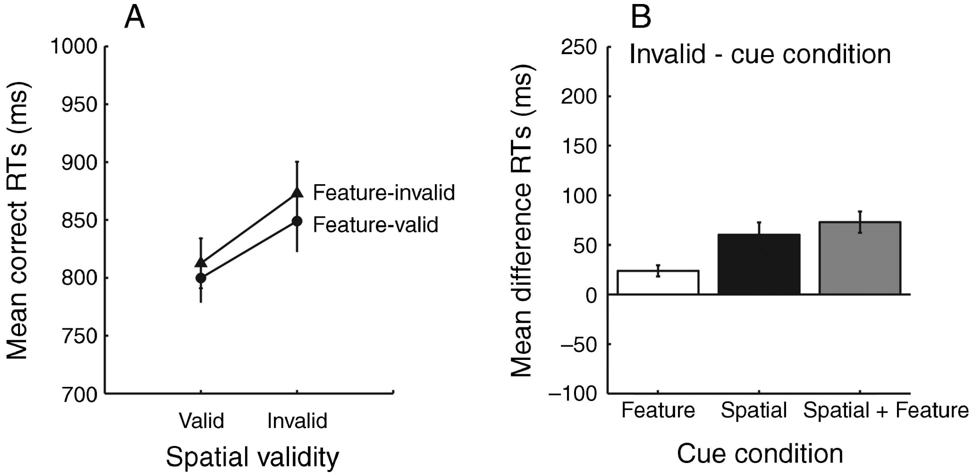

Figure 3. (A) Mean correct response times (RTs) for Experiment 3. Experiment 3 replicates the results of Experiment 2 using identical displays but changing the task from a detection task to a discrimination task. Mean correct reaction times suggested that response times to spatial and feature valid cues were faster than response times to spatial and feature invalid cues. Interestingly, there was no significant interaction between spatial and feature validity, indicating that the spatial and feature validity benefits did not interact to produce a multiplicative reduction in reaction times. Error bars represent 1 *SEM*. (B) Mean correct difference reaction times (RTs) for Experiment 3. Cue benefits represent the difference between RTs from both spatially and featurally invalid trials and the three valid cue types (spatially valid, featurally valid, and both spatially and featurally valid). Error bars represent 1 *SEM*.

Again and as expected from Experiment 1, each of the valid cue conditions resulted in faster reaction times than the spatial and feature invalid condition (*spatial valid vs. spatial and feature invalid:* $t(17) = 4.86$, corrected $p < 0.001$; *feature valid vs. spatial and feature invalid:* $t(17) = 4.22$, corrected $p < 0.01$; *spatial–feature valid vs. spatial and feature invalid:* $t(17) = 6.80$, corrected $p < 0.001$). Furthermore, while a spatially valid cue produced a larger RT benefit than a featurally valid cue ($t(17) = 3.06$, corrected $p = 0.014$) and although there were no significant interactions found in the repeated-measures ANOVA, RTs for spatial–feature valid cues were no longer significantly faster than RTs for spatial valid alone ($t(17) = 1.48$, corrected $p = 0.16$) in Experiment 3.

Analysis of accuracies indicated no evidence of a speed–accuracy trade-off in this experiment. Average accuracy rates were lower than in Experiment 1's detection task but still very high (spatial and feature invalid = 87.71%, *SEM* = 5.15%; spatial valid = 89.76%, *SEM* = 5.51%; feature valid = 88.78%, *SEM* = 4.76%; spatial and feature valid = 89.00%, *SEM* = 7.28%). A one-way ANOVA revealed that accuracy did not vary significantly across the four conditions, $F(3,51) = 1.28$, $p = 0.29$, just as in Experiment 1.

General discussion

In Experiment 1, we found that uninformative and task-irrelevant exogenous cues that precede a visual search array can produce reaction time benefits for targets that share a color with the cue, even if the cue and the target appear at remote spatial locations. Since the cues are completely non-predictive about the location of the target in the subsequent search task, we hypothesized that the search benefit was produced by a rapid and reflexive spread of feature-based attention that resulted in enhanced processing for items that shared a color with the exogenous cue. Interestingly, though the effect seemed to be independent of more well-known mechanisms of spatial exogenous cuing, it did not reliably appear when search arrays of only three items were used rather than six. Furthermore, in Experiment 2, we found that cues presented at fixation rather than at target locations did not produce any discernable feature-based benefit, suggesting that the exogenous cues needed to capture attention in order for the effect to manifest. Finally, in Experiment 3, we found that the effect appears not only in detection judgments but also when subjects have to discriminate some characteristic of the target oval, leading us to conclude that our results constitute evidence for a previously unreported and unknown type of visual attention: reflexive feature-based attention.

To date, several studies have examined how color features can influence contingent attentional capture by drawing attention to color-salient items in the visual field (Folk, Remington, & Johnston, 1992; Folk, Remington, & Wright, 1994; Theeuwes, 1994; Theeuwes & Burger, 1998). However, in all of these studies, color marked either a highly salient target or was a task-relevant feature in the subsequent visual search. In contrast, in our method, neither the color nor the spatial location of the cue provided any predictive information as to the color or location of the target oval. Furthermore, targets in our search task were defined by an attribute (shape or oval orientation) completely orthogonal to those of the cues. In other words, even if our participants had been able to adjust their search strategy or attentional allocation in response to the color of the cue, they would have gained no reaction time or accuracy benefit in detecting or discriminating the target over the whole course of the experiment. To our knowledge, the current study is the first to demonstrate a reflexive feature-based benefit on visual search times when the color feature is completely irrelevant to the task.

In Experiment 2, no significant effect was found when the exogenous feature-based cue was presented at fixation. Given that effects were found in Experiment 1 when feature-based cues were inextricably linked with non-foveal spatial cues, but not in Experiment 2, it is possible that attentional capture to the spatial cue is a necessary condition for the exogenous feature-based attention effect. There is conflicting research regarding how the size and shape of the attentional window modulates attentional capture (Belopolsky & Theeuwes, 2010; Belopolsky, Zwaan, Theeuwes, & Kramer, 2007; Eriksen & St James, 1986) or whether top-down strategies can override attentional capture (Leber & Egeth, 2006). Other experiments that have looked at attentional capture have shown puzzling results with foveal cues; for example, Folk et al. (1992) presented abrupt onset foveal cues and expected to find significant costs due to the cue on search for a subsequent target compared to a neutral cuing condition, but no costs were found, suggesting that attention was not captured or affected by the foveal cue. Because visual cues must inevitably contain spatial information, Folk et al. also suggested that foveal cues can be extremely problematic given that they always signal a non-target location and thus are conceptually 100% invalid cues. In light of this, presenting our exogenous feature-based cues at the fovea where fixation and attention was presumably already allocated prior to the trial start could have reduced the size of the feature-based effect if attentional capture plays a modulating role in the effects of exogenous feature-based attention.

Is it possible that our results are not due to a reflexive spread of feature-based attention but are instead due to some form of color priming? Research in the past has suggested that repeated color or spatial location priming across trials does modulate performance (Kristjánsson, Ingvarsdóttir, & Teitsdóttir, 2008; Kristjánsson, Vuilleumier, Malhotra, Husain, & Driver, 2005) and color priming on one trial can leave a memory trace lasting between 6 and 17 s that can affect performance on subsequent trials (Goolsby & Suzuki, 2001). However, given the relatively long time scale such effects are thought to operate over, it seems unlikely that the same mechanism is responsible for the extremely rapid within-trial effects we observe here. Furthermore, in many of these previous studies, the color primes had relative contingencies with the eventual targets such that the color primes were often informative about the eventual target location. In the current paradigm, the color of the cues varied nearly randomly from trial to trial, reducing the effects of any lingering memory traces. Furthermore, the color and spatial features of the cue in our experiments were not predictive of the targets in any way but were completely irrelevant to the task. Taken together, this suggests that the current results were truly reflexive in nature and not the result of cross-trial priming.

Our results have three important implications on models of attention and how different forms of attention interact. First, the results extend previous literature on feature-based attention by showing for the first time that feature-based attentional effects can be produced by a reflexive mechanism that operates on a very short time scale and appears to require no internal guidance. Second, these effects can be elicited by exogenous cues that are irrelevant and uninformative to the task at hand. Third, these results also suggest that exogenous feature-based and spatial attention effects operate independently but can be combined to produce an additive benefit in a search task. Results from the current paper support recent evidence that spatial- and feature-based aspects of the attention control system can be derived independently and interact additively (Hayden & Gallant, 2009; Patzwahl & Treue, 2009). In both Experiments 1 and 3, spatial attention effects appeared to be greater than feature-based effects. Recent research has suggested that the speed of feature-based attention is relatively slow compared to endogenous spatial attention (Huang, 2010; Liu, Stevens, & Carrasco, 2007). With this in mind, it may be that in feature valid trials attention is first captured and drawn to the spatial location of the cue. Consequently, feature-based attention spreads throughout the visual field after the processing of the irrelevant color feature on the spatial cue. Because attention is always initially captured to a distractor location in feature valid trials but captured directly to the correct target location in spatial valid trials, this could potentially explain why feature-based effects always appear smaller than spatial-based effects. Future research will have to be conducted to investigate whether attentional capture or the size and shape of the attentional window modulate effects of feature-based attention from exogenous cues.

Acknowledgments

Authors JYL and BHW contributed equally to this project. We thank two anonymous reviewers and the editor for their helpful comments and Fabian Santana for his help with data collection. JYL was supported as a Howard Hughes Medical Institute International Student Research Fellow. For BHW, this material is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0718124. This work was supported by National Institutes of Health (NIH) Grant EY12925 to GMB.

Commercial relationships: none.

Corresponding author: Jeffrey Y. Lin.
Email: jytlin@u.washington.edu.
Address: University of Washington, Guthrie Hall, 4000 15th Ave. NE, Box 351525, Seattle, WA 98195, USA.

References

Belopolsky, A. V., & Theeuwes, J. (2010). No capture outside the attentional window. *Vision Research, 50,* 2543–2550. [PubMed]

Belopolsky, A. V., Zwaan, L., Theeuwes, J., & Kramer, A. F. (2007). The size of an attentional window modulates attentional capture by color singletons. *Psychonomic Bulletin & Review, 14,* 934–938. [PubMed]

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Reviews in Neuroscience, 33,* 1–21. [PubMed]

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature, 408,* 196–199. [PubMed]

Boynton, G. M., Ciaramitaro, V. M., & Arman, A. C. (2006). Effects of feature-based attention on the motion aftereffect at remote locations. *Vision Research, 46,* 2968–2976. [PubMed]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51,* 1484–1525. [PubMed]

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. [PubMed]

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Reviews in Neuroscience, 18,* 193–222. [PubMed]

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269–297. [PubMed]

Eriksen, C. W., & St James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40,* 225–240. [PubMed]

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030–1044. [PubMed]

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 317–329. [PubMed]

Giordano, A. M., McElree, B., & Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed–accuracy trade-off analysis. *Journal of Vision, 9*(3):30, 1–10, http://www.journalofvision.org/content/9/3/30, doi:10.1167/9.3.30. [PubMed] [Article]

Goolsby, B. A., & Suzuki, S. (2001). Understanding priming of color-singleton search: Roles of attention at encoding and "retrieval". *Perception & Psychophysics, 63,* 929–944. [PubMed]

Hayden, B. Y., & Gallant, J. L. (2009). Combined effects of spatial and feature-based attention on responses of V4 neurons. *Vision Research, 49,* 1182–1187. [PubMed]

Huang, L. (2010). The speed of feature-based attention: Attentional advantage is slow, but selection is fast. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1382–1390. [PubMed]

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology, 37,* 43–75. [Article]

Kinchla, R. A. (1992). Attention. *Annual Review of Psychology, 43,* 711–742. [Article]

Kristjánsson, A., Ingvarsdóttir, A., & Teitsdóttir, U. D. (2008). Object- and feature-based priming in visual search. *Psychonomic Bulletin & Review, 15,* 378–384. [PubMed]

Kristjánsson, A., Vuilleumier, P., Malhotra, P., Husain, M., & Driver, J. (2005). Priming of color and position during visual search in unilateral spatial neglect. *Journal of Cognitive Neuroscience, 17,* 859–873. [PubMed]

Leber, A. B., & Egeth, H. E. (2006). It's under control: Top-down search strategies can override attentional capture. *Psychonomic Bulletin & Review, 13,* 132–138. [PubMed]

Lin, J. Y., Pype, A. D., Murray, S. O., & Boynton, G. M. (2010). Enhanced memory for scenes presented at behaviorally relevant points in time. *PLoS Biology, 8,* e1000337. [PubMed]

Liu, T., & Mance, I. (2011). Constant spread of feature-based attention across the visual field. *Vision Research, 51,* 26–33. [PubMed]

Liu, T., Stevens, S. T., & Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research, 47,* 108–113. [PubMed]

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103,* 603–649. [PubMed]

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences, 29,* 317–322. [PubMed]

McAdams, C. J., & Maunsell, J. H. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. *Journal of Neurophysiology, 83,* 1751–1755. [PubMed]

Nobre, A. C., Allison, T., & McCarthy, G. (1998). Modulation of human extrastriate visual processing by selective attention to colours and words. *Brain, 121,* 1357–1368. [PubMed]

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology, 52,* 629–651. [PubMed]

Patzwahl, D. R., & Treue, S. (2009). Combining spatial and feature-based attention within the receptive field of MT neurons. *Vision Research, 49,* 1188–1193. [PubMed]

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3–25. [PubMed]

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Reviews in Neuroscience, 13,* 25–42. [PubMed]

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology, 58,* 1–23. [PubMed]

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5,* 631–632. [PubMed]

Sàenz, M., Buracas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research, 43,* 629–637. [PubMed]

Schreij, D., Owens, C., & Theeuwes, J. (2008). Abrupt onsets capture attention independent of top-down control settings. *Perception & Psychophysics, 70,* 208–218. [PubMed]

Schreij, D., Theeuwes, J., & Olivers, C. N. (2010). Abrupt onsets capture attention independent of top-down control settings II: Additivity is no evidence for filtering. *Attention, Perception & Psychophysics, 72,* 672–682. [PubMed]

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron, 55,* 301–312. [PubMed]

Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition, 115,* 118–132. [PubMed]

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 799–806. [PubMed]

Theeuwes, J., & Burger, R. (1998). Attentional control during visual search: The effect of irrelevant singletons. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1342–1353. [PubMed]

Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399,* 575–579. [PubMed]

White, A. L., & Carrasco, M. (2011). Feature-based attention involuntarily and simultaneously improves visual performance across locations. *Journal of Vision, 11*(6):15, 1–10, http://www.journalofvision.org/content/11/6/15, doi:10.1167/11.6.15. [PubMed] [Article]

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 676–681. [PubMed]

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 661–676. [PubMed]

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621. [PubMed]

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 121–134. [PubMed]

VITA

Jeffrey Yu-Ting Lin was born in Kaohsiung, Taiwan. He earned his Bachelors at the University of British Columbia and a Master of Science in Psychology at the University of Washington. In 2012, he earned a Doctor of Psychology at the University of Washington.