Efficient Training Of Visual Search Via Attentional Highlighting

Michael C. Mozer

Institute of Cognitive Science, University of Colorado

Harold Pashler

Department of Psychology, UC San Diego

Robert Lindsey

Department of Computer Science, University of Colorado

Jason Jones

Department of Psychology, UC San Diego

Corresponding author: Michael Mozer

Email: `mozer@colorado.edu`

Phone: +1-303-517-2777

Mailing address: Institute of Cognitive Science, University of Colorado, Boulder CO 80309–0344

keywords: visual search, expertise, training, attentional control

Abstract

Human activities often involve interaction with dynamic or complex visual environments, e.g., driving an automobile or inspecting radiographic images. Knowing where to look and what features to look for in a given context is critical to expertise, but this knowledge is typically acquired by effortful and inefficient trial-and-error learning. We ask whether training of endogenous attentional control can be improved by *highlighting* regions of interest in a display. Participants searched displays whose global configurations predicted the location of a target. Training reduced response latencies—indicating learning of contingencies—whether during training participants effortfully searched for targets or were directed to target locations via highlighting (brief flashes). However, highlighting was effective only when the training task required no response beyond a saccade to the target. By halving training time and demanding less effort, highlighting shows practical potential and also provides insight into the interplay between exogenous and endogenous attentional control.

Introduction

Human activities often involve interaction with a complex visual environment, e.g., driving a car, piloting a plane, controlling air traffic, screening baggage, matching fingerprints, analyzing satellite images for military intelligence, inspecting x-ray and MRI images for abnormalities. Expertise in such activities comes from experience over a time period of years (Newell & Rosenbloom, 1981). Becoming an expert involves two distinct abilities: identifying features and locations in the environment that are task relevant at a given moment in a given context, and determining the appropriate action to take. Acquiring these two abilities poses a chicken-and-egg problem: the appropriate response cannot be determined until one knows what is relevant, but what is relevant depends on it providing reliable information about the task-appropriate response.

Novices often master visual skills by trial and error experimentation. This contrasts with learning in declarative domains such foreign language vocabulary, where it would be extremely inefficient to teach a student by trial and error (e.g., presenting a French word and having them guess possible English equivalents until they happen on the right one.) However, explicit instruction in complex visual domains is difficult. Efforts to design tutoring systems for interpretation of complex visual images, e.g., mammograms (Azevedo & Lajoie, 1998; Guliato et al., 2009), have focused on cognitive training (instruction and descriptions) and software environments (e.g., tools for highlighting regions of interest) not on the dynamics of gaze. Experts might be used to elicit visual features that are relevant for decision making, and the list of features be provided to the novice (Getty, Pickett, D'Orsi, & Swets, 1988). One could ask an expert to stand over the shoulder of a novice as they, say, tried to control a flight simulator, and the expert could provide guidance such as, "Check your altimeter now." However, even if it is feasible for a vigilant expert to monitor training, other issues abound: experts often lack the ability to self-report and verbalize their knowledge (e.g., Feldon, 2007), rapid fire perceptuomotor decision making may preclude verbal communication, and such interruptions may even prove harmful.

Our long-term goal is to design training systems that utilize the implicit knowledge of experts to assist in the efficient training of novices, in the same spirit as Tomlinson, Howe, and Love (2009).[1] Rather than ask experts to provide explicit guidance, we aim to leverage expert performance in a novel perceptual learning paradigm involving the following steps.

1. Record the fixation sequences of experts as they perform a complex visual task in front of a computer monitor.
2. Using machine learning techniques and with computational models of human spatial attention, construct a discriminative model that predicts locations that experts would be likely to inspect and novices would be unlikely to inspect given the current visual context and task state.
3. Place the novice in the visual environment and have them perform the task.
4. In parallel with the novice performing the task, use the discriminative model to dynamically suggest regions of interest to the novice at each instant.
5. Highlight the regions of interest via saliency manipulations of the visual display. These manipulations can be as simple as a brightness or contrast modulation that would increase the probability that a novice would be drawn to a location purely by exogenous guidance of attention. Or they might be more complex and involve modulation of feature discriminability (Whitehill & Movellan, 2012).

Although effective guidance of spatial attention is straightforward, the experimental literature provides no strong support for the notion that directing a learner's attention to the appropriate location will facilitate training. Nonetheless, cueing has been shown to influence other aspects of cognition, specifically problem solving, choice, and detection. Subtle perceptual cues that direct participants' gaze or induce covert attentional shifts improve the success rate in solving a diagram-based insight problem (the tumor-and-lasers

---

[1] Tomlinson et al. (2009) studied a video game in which players could select one of eight different status information formats for an on-screen display. Using a model of expert selection as a training companion, novices provided with contextually relevant information converged more rapidly on expert-like behavior.

radiation problem) (Grant & Spivey, 2003; Thomas & Lleras, 2009). Directing attention to an object can also bias simple choice (Armel, Beaumel, & Rangel, 2008). And novice performance on detecting anomalies in a chest x-ray improves when shown expert radiographer's scan paths on the same image (Litchfield, Ball, Donovan, Manning, & Crawford, 2010). However, improvements are also observed when novices were shown scan paths of other novices performing the detection task, suggesting perhaps a wisdom-of-crowds effect wherein an individual benefits by scaffolding their decisions on top of those of another. Some negative results have been reported as well. Advising individuals where to look does not always lead to benefits in comprehension (Kriz & Hegarty, 2007; Koning, Tabbers, Rikers, & Paas, 2010) or performance (Donovan, Manning, Phillips, Highman, & Crawford, 2005), possibly because instructions to fixate may hinder processing of fixated information (Dewhurst & Crundall, 2008).

The goal of the present work is to determine whether providing a novice with task- and context-appropriate indication of where to look—which we'll refer to as *highlighted training*—will enhance learning to search. Optimistically, because highlighting helps novices avoid trial-and-error behavior, they might more effectively learn where to look in a given context. From a more neutral perspective, associations between display contexts and saccadic responses may be strengthened simply by performing them (Guthrie, 1952), in which case one would expect little difference between highlighted training and trial-and-error behavior. And finally, from a pessimistic perspective, highlighted training may simply provide the novice with a crutch that allows them to avoid learning altogether, thus encouraging a sort of lazy behavior. The idea that lazy learning may be ineffective is a key tenet of the desirable difficulty doctrine of Bjork and colleagues (e.g., Schmidt & Bjork, 1992). According to these researchers, learning crutches often facilitate performance during training, while actually reducing the amount that people learn—as assessed when the crutches are withdrawn.

We explore the impact of highlighted training in visual search. In Experiments 1 and 2, participants are asked to search for a sideways T among L's and to report whether the T lies on its left or right side (Figure 1). Adopting the classic contextual cueing paradigm (Chun & Jiang, 1998), some configurations of

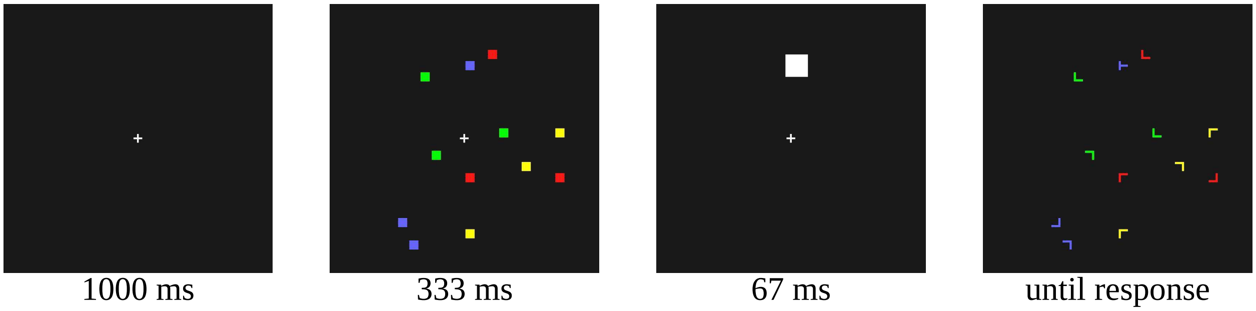


Figure 1: Sequence of events within a trial, Experiments 1 and 2. The highlight (flash) that guides attention is included in only some training conditions. When no highlight is present, the brief frame contains only a fixation cross.

the distractors are repeated over the course of the experiment, with the target appearing in a fixed position relative to the configuration. Repetition allows participants to discover and exploit the contingency between visual context and the target location. The degree to which participants have learned to guide attention in a context-appropriate manner is reflected in the speed up for these *repeated* configurations relative to *novel* configurations that have not previously been presented in the course of the experiment.

By definition, visual search requires participants to locate the target among distractors. Our interest is in evaluating the effectiveness of training in which participants need not perform effortful search. In a highlighted-training condition, we present a brief bright *flash* prior to the onset of the stimulus array that directs attention to the target. This flash is included in the *training phase*, and is removed in a *testing phase*, during which the effectiveness of highlighted training can be evaluated relative to traditional effortful-search training. (We use the term flash instead of cue to avoid confusion with the ’cue’ in contextual cueing.)

Experiment 1

Method

On each trial, a stimulus array was presented containing a target and 11 L distractors in random orientations, with an equal number of elements colored red, blue, yellow, and green.

Three conditions were compared: repeated, novel, and guided-repeated. The latter condition involves repeated configurations preceded by a flash which highlighted the target in the training phase. Each condition consisted of eight stimulus arrays presented a total of 32 times, with order permuted randomly within a block. For the repeated and guided-repeated conditions, the arrays were identical across blocks except the target direction alternated between pointing left and right across blocks. For the novel condition, the target color and location were constant across blocks but the target direction and distractor arrangement changed from block to block. This ensured that the context-independent predictability of target location and color was comparable across conditions.

In pilot studies, we intermixed the three conditions but found that participants came to depend on the flashes, which appeared to cause anticipatory confusion on the 2/3 of trials in which no flash occurred. Consequently, we ran a variant of Experiment 1 in which the three conditions were *blocked*, ensuring that flashes were uniformly present or absent over a training sequence. However, this design put the guided-repeated condition at a disadvantage in the testing phase because at the start of the testing phase, participants faced the sudden removal of the highlights (versus the other two conditions which had no highlighting). Consequently, we also conducted a version of Experiment 1 in which the three conditions were *mixed*, and every trial in both training and test phases included a flash. The flashes were 100% valid for the guided-repeated items in the training phase, but were 100% invalid otherwise. Nonetheless, during the training phase, the flash validity was higher than chance (33% valid on average across the three intermixed conditions, versus 8% by chance). The mixed design avoided the sudden disappearance of flashes in the test phase, but required invalid flashes during the training phase.

For data analysis, every four consecutive blocks were grouped to form an *epoch*. The training and test phases consisted of six and two epochs, respectively. The blocked and mixed designs were each run (between subjects) with 24 UCSD undergraduates. Participants were rejected if their error rate exceeded 3/32 (about 10%) in any epoch in any condition. Five rejected participants were replaced in blocked design, to ensure block order was counterbalanced. Four participants were rejected in the mixed design but were not replaced.

Results

Figure 2 shows results separately for the blocked and mixed designs. Because no meaningful interactions are found involving this factor, all analyses collapse across the design. Error rates are low in all conditions ($< 1\%$) and no statistically significant effects are observed, suggesting that RTs can be analyzed without concern about speed-accuracy trade offs.

The contextual cueing phenomenon (Chun & Jiang, 1998) is replicated: The repeated condition is

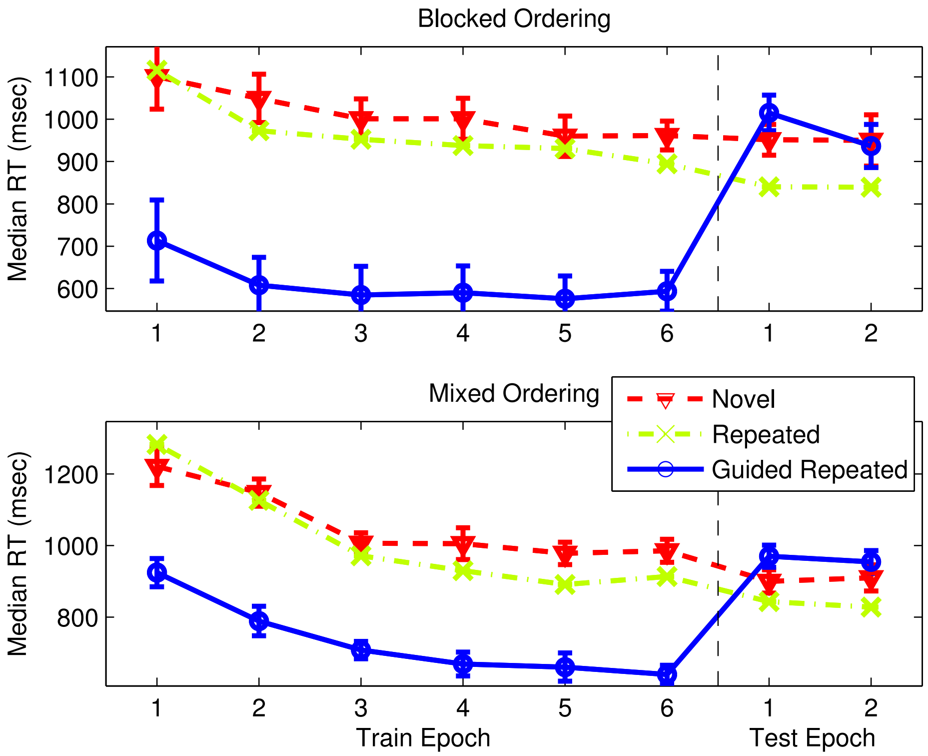


Figure 2: Experiment 1, reaction times (RTs) for the blocked and mixed designs (upper and lower figures, respectively). RT is graphed for the six training epochs followed by the two test epochs. The target is highlighted during training in the guided-repeated condition, but not in the other two conditions. Contextual contingencies are predictable in the repeated and guided-repeated conditions, but not in the novel condition. Mean RTs are computed across participants using each individual's median RT, and error bars indicate +/- 1 SEM based on the paired differences between the repeated condition and each of the other two conditions.

faster than the novel condition ($F(1,42) = 4.47, p = .040$) and repetition interacts with epoch ($F(7,294) = 2.80, p = .008$), reflecting learning about the repeated displays through experience. There is also a main effect of epoch in these two conditions ($F(7,294) = 37.86, p < .001$) reflecting learning to search and respond appropriately in the task.

The highlighting manipulation was clearly effective: during the training phase, RTs are up to 50% faster in the guided-repeated condition than in the repeated condition.

Although participants discover contextual contingencies when they effortfully search for targets, there is no evidence that learning takes place when attention is guided to the target during training. When highlighting is removed in the two test epochs, guided-repeated trials are slower than repeated trials ($F(1,42) = 23.57, p < .001$), and appear to be slower than novel trials, although that difference is not reliable ($F(1,42) = 2.45, p = .125$).

These discouraging results suggest the possibility that guidance per se interferes with learning of contextual contingencies. However, consider an alternative hypothesis that stems from a reinforcement learning perspective. On each trial, participants perform a sequence of actions that lead to a successful or unsuccessful conclusion to the trial. These actions consists of one or more saccades and/or internal shifts of attention, perceptual processing, and ultimately the initiation of a discriminative choice response (reporting whether the target points to the left or right). Following receipt of punishment—an error tone and delay—or reward—the trial ends and the participant is closer to finishing the experiment—reinforcement learning processes face a problem of *temporal credit assignment* (Sutton & Barto, 1988) to determine the action or actions in the sequence responsible for the trial outcome. With discounting, actions at the end of the sequence are given more credit or blame than actions earlier in the sequence. In Experiment 1, primary responsibility would be assigned to the target discrimination and response (left or right T), and weaker responsibility to the prior saccade or attention shift to the target. It seems plausible to suppose that credit assignment is modulated by whether actions are exogenously or endogenously generated: the brain gives itself less credit for an attention shift initiated by the flash than one internally initiated. If so, perhaps the benefit for highlighted training is abolished by the *combination* of temporal discounting and the further reduced credit for exogenously triggered behaviors.

By this hypothesis, highlighted training would increase in effectiveness if participants merely had to attend to the target but were not required to respond. Using stimuli identical to those in Experiment 1, highlighting occurred on all trials of the training phase of Experiment 2. On the large majority of trials, the target location was highlighted. On occasional trials, the target was absent, and one of the distractors was highlighted instead. Participants were instructed to press a response key only on target-absent trials. This occasional response ensured that vigilance was maintained during training. During a subsequent test phase, participants performed the left versus right T discrimination task.

Experiment 2

Method

Displays in Experiment 2 were like those used in Experiment 1. The training phase consisted of a series of blocks each containing 12 *guided-repeated* displays and 12 *guided-novel* displays, intermixed and randomized within a block. On one trial per condition per block, the target was replaced by a distractor. On all trials, a flash was presented indicating the location of the target (or where the target would have been). A trial ended either when the participant pressed the 'target absent' key or when one second had passed. Error feedback was provided. Training trials were quick and relatively effort free: target-present trials lasted 2.4 seconds, target-absent trials less.

An *untrained-repeated* condition with 12 new repeated items was added during the test phase, which thus consisted of 36 trials per block. Results were analyzed by grouping three blocks into an *epoch*, with a total of 8 training epochs and 4 test epochs.

Sixty UCSD undergraduates participated for course credit.[2] Six participants were rejected because their

---

[2] We initially planned to run 24 subjects, as in Experiment 1, but the additional subjects were run due to spare lab capacity. The

error rate during training exceeded 1/12, which is the rate that would be obtained by not responding to target-absent trials. Five of the remaining 54 participants were rejected for having an error rate in one epoch of one condition that exceeded 4/36 (about 10%).

Results

Figure 3 shows RTs in the three conditions during the test phase. Participants improve over the test phase ($F(3,144) = 28.7, p < .001$) with no reliable interaction of epoch with condition ($F(6,288) = 1.23, p = .289$). Performing a paired comparison, the guided-repeated condition was reliably faster than the guided-movel condition ($F(1,48) = 13.89, p = .001$; interaction with epoch: $F(3,144) < 1$), indicating that highlighted training obtained learning about contextual contingencies in Experiment 2, in contrast to Experiment 1. The key difference between the two experiments is that Experiment 1 demanded a discriminative response on each training trial (whether target points to the left or the right) whereas Experiment 2 did not.

Learning was fairly robust and persistent, as evidenced by the fact that the advantage for the guided-repeated condition over the untrained-repeated condition persisted throughout the test phase (main effect: $F(1,48) = 4.06, p = .05$; interaction with epoch: $F(3,144) < 1$). Although one might expect the untrained-repeated displays to eventually catch up with the guided-repeated displays, they did not do so in 12 test-phase presentations of each display.

Having established that highlighted training is effective when most training trials demand no response (Experiment 2), in contrast to when each training trial requires a discriminative response (Experiment 1), we conducted a final experiment to explore the consequences of eliminating the demand to perform any task or make any response during training beyond a saccade to the target.

---

same results were obtained with the first 24, and in fact, the results were a bit more robust, perhaps because the additional 36 came at the quarter's end.

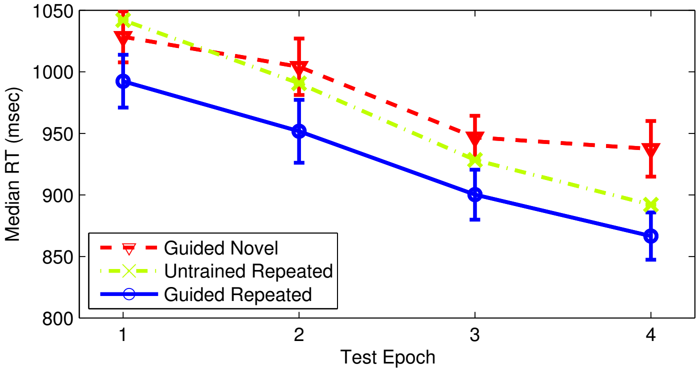


Figure 3: Experiment 2, test epoch RTs. Prior to test, participants were trained (with highlighting) in the the guided-repeated and guided-novel conditions, but not in the untrained-repeated condition. Contextual contingencies are predictable in the guided-repeated and untrained-repeated conditions, but not in the guided novel condition. Mean RTs are computed across participants using each individual's median RT, and error bars indicate +/- 1 SEM based on the paired differences between the untrained-repeated condition and the other two conditions.

Experiment 3

In Experiment 3, participants were monitored by an eye tracker and each trial terminated when a saccade was made to the target location. Because no discriminative response to the target was required, we eliminated the target, or rather, made the target *invisible*. Participants were shown distinct brightly colored *martian cloud* contexts (Figure 4) and were instructed to search for an invisible ship hidden in one quadrant. In the *repeated* condition, the target quadrant associated with a context was fixed across blocks; in the *novel* condition, the target quadrant changed each block.[3] A trial ended with a saccade to the outer portion of the quadrant. (Kunar, Flusberg, & Wolfe, 2006) previously demonstrated contextual cueing based on background color and texture; the martian-cloud contexts provide both.

Method

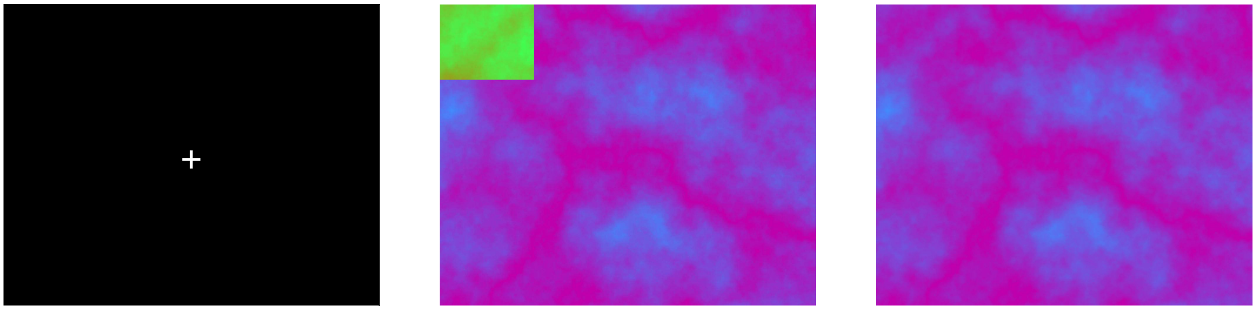
The structure and organization of this experiment was identical to that of Experiment 1. The blocked versus mixed design was included as a between-subjects factor, with novel, repeated, and guided-repeated conditions run within subjects. Twelve participants were run in each design, with condition order counterbalanced in the blocked design. As in Experiment 1, there were 6 epochs of training followed by 2 epochs of testing, each epoch composed of 4 blocks and each condition consisting of 8 contexts per block. Target quadrant was counterbalanced within blocks and conditions. For the mixed design, the invalid flash quadrant was counterbalanced within blocks in novel and repeated conditions.

Each trial ended only when participants fixated in the outer 25% of the target quadrant. Consequently, errors were not possible. Eye tracking was performed with a Tobii 1750 with Clearview Software version 2.7.1. Several participants were not run to completion due to problems with the eye tracker and were replaced.

Context intensity patterns were produced by $1/f^{1.75}$ noise, and the two colors of each context were cho-

---

[3] 'Repeated' and 'novel' might better be termed 'predictive' and 'nonpredictive' contexts in this Experiment, but we kept the same nomenclature for consistency with Experiments 1 and 2.



until cross fixated for 500 msec

50 msec

until target quadrant fixated for 50 msec

Figure 4: Sequence of events within a trial, Experiment 3. A fixation cross is presented until the participant fixates within 10% of the center of the screen for 500 msec. To guide attention, some trials include a highlight (a 50 msec flash) in the outer corner of one quadrant. The flash is a negative of the context pattern to ensure high discriminability from the context. The context then appears until the participant fixates on outer corner of the target quadrant for 50 msec. RT is measured from the onset of the flash or from the onset of the context alone if no flash is presented.

sen to maximize discriminability among the alternative contexts. Sixty-four distinct contexts were generated and eight of these were randomly drawn for each condition and participant.

Results

No interactions were found involving the design (blocked versus mixed); consequently, all analyses and figures collapse across this factor. The upper graph in Figure 5 shows the RTs by epoch and condition. The highlighting manipulation was successful: during the training phase, the RT in the guided-repeated condition was about half that of the repeated condition. The contextual cueing manipulation also appeared to be successful, with main effects of novel versus repeated ($F(1,22) = 24.35, p < .001$) and epoch ($F(7,154) = 14.78, p < .001$) during the training phase but no reliable interaction ($F(5,110) = 0.34$). The absence of an interaction may reflect the fact that the training effect occurred in the first few blocks and therefore is buried within the first epoch.

Crucially, a robust effect of highlighted training is observed during the test phase. As in Experiment 2, the guided-repeated condition is faster than the novel ($F(1,22) = 6.49, p = .018$) and not reliably different than the repeated ($F(1,22) = 0.38$).

Although contextual cueing by global stimulus features unarguably speeds up search, Kunar et al. (2006) found that it produces at best a weak improvement in search efficiency, i.e., the processing time required per element in multielement displays. They thus claimed that this form of learning of contextual contingencies does not facilitate attentional guidance. Experiment 3 provides two arguments against this conclusion. First, because no target is physically present, learning cannot be attributed to an improvement in detecting a target in a familiar context; participants must be learning contextually appropriate behavior. Second, and perhaps stronger, search efficiency can be directly observed in Experiment 3 via the saccade sequence. The number of fixations per trial (lower graph, Figure 5) parallels the response times per trial (upper graph). Most relevant, fewer fixations are made during the test phase in the guided-repeated condition than in the novel ($F(1,22) = 5.31, p = .031$). Similar patterns were observed with the total number of quadrants visited and

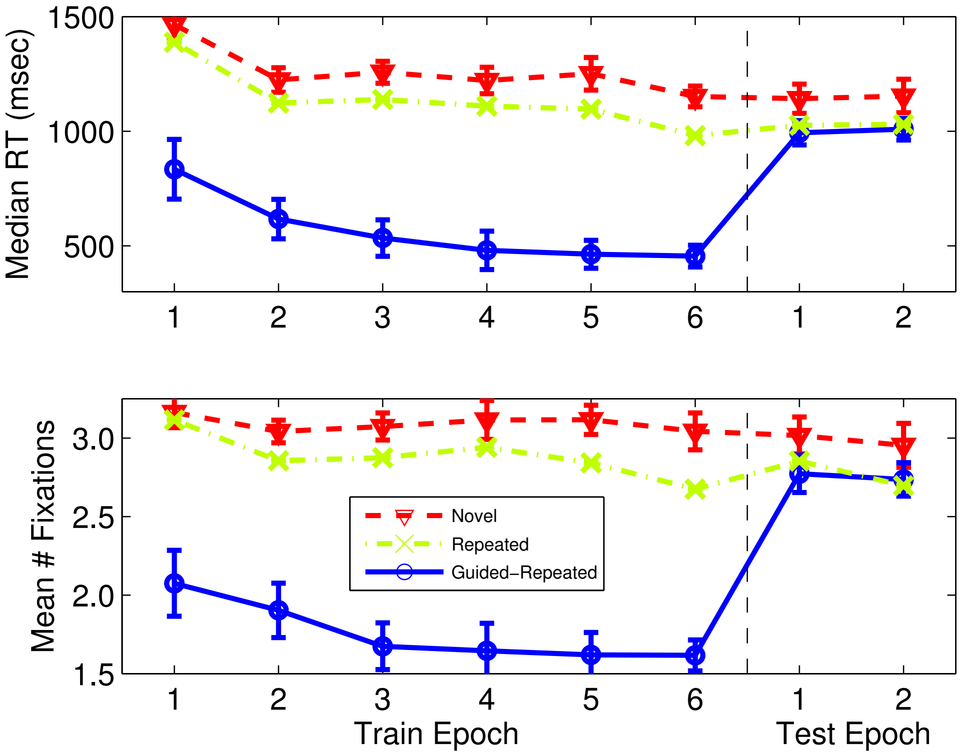


Figure 5: Experiment 3, RTs (upper figure) and number of fixations per trial (lower figure). In this experiment, the trial ended successfully when the participant fixated near a corner of the target quadrant. No further response was necessary. Mean RTs are computed across participants using each individual's median RT; mean number of fixations is computed using each individual's mean, with outlier trials discarded (the number of fixations is more than 3 standard deviations above the mean). Error bars indicate +/- 1 SEM based on the paired differences between the untrained-repeated condition and the other two conditions.

the bias participants had to return to the target quadrant from the previous trial.

Discussion

We have demonstrated that highilighting—guiding attention to a task-relevant location in the visual field—can facilitate learning of contextual dependencies. Highlighted training produces the same degree of learning as effortful search, but cuts the time per trial in half. Because we examined performance after a fixed number of training trials, we cannot determine whether the effectiveness of training with and without highlighting differs on a per trial basis. However, we can conclude that the total time required to achieve a certain level of performance can be significantly reduced with highlighting. Moreover, highlighted training strikes participants as demanding little exertion, concentration, or volitional control, and therefore may be particularly effective when motivation levels cannot be sustained.

The failure of highlighted training in Experiment 1 suggests that performing even a simple task following the attention shift can interfere with the learning of contextual contingencies. Consequently, a two-phase training procedure seems warranted: first an exploratory phase guides the learner to relevant location in a variety of contexts via highlighting, then an associative phase trains appropriate responses at those locations.

The theoretical implications of this work concern the interplay between exogenous and endogenous attentional control. Although these two forms of guidance are distinguished in most theoretical treatments of attention (see Wilder, Mozer, & Wickens, 2011 for a review), our results suggest that they are not independent, and exogenous cues—the flashes in our experiments—can serve as a training signal for context-dependent endogenous control. However, this process of internalization occurs only when visual exploration is isolated from other cognitive components of the task.

From a practical perspective, our results encourage exploration of the following scenario for, say, training student drivers. The student might view a video taken from the driver's perspective of an automobile as it is driven on the road. Highlights regularly direct the student's gaze to critical elements of the video stream—e.g., a car signaling to change lanes or enter the road, a pedestrian standing between parked cars—and between different views of the road—from the front windshield to the rear- and side-view mirrors. Although clearly such exposure cannot be more effective than actual driving experience, it comes without the cost of putting a novice behind the wheel. Moreover, training may be further concentrated by playing the video at faster-than-real-time rates. Further, pretraining attention may serve to amplify the effects actual practice on the skill.

We conjecture that the potential benefits of highlighted training are greater than than the present studies suggest, for two reasons. First, in naturalistic tasks, similar contexts demand similar responses, yet the contexts in our experiments did not permit generalization from one to another. Consequently, in naturalistic tasks, transfer benefits may increase the value of training. Second, some evidence indicates that the facilitative effects of visual cues diminishes with task experience (Wright & Richard, 1999). If so, decreasing the salience of the highlights over time may increase the effectiveness of highlilghted training.

Highlighting might prove useful for guiding attention not only to relevant locations, but to other stimulus dimensions. For example, Pashler and Mozer (2012) show that exaggerating values on a stimulus dimension relevant to classification can enhance training when stimuli are high dimensional and the critical classification dimension doesn't inherently stand out to the learner. We hope that attentional highlighting points us in the direction of identifying a broader family of techniques for enhancing learning.

Acknowledgments

We thank Noriko Coburn and Tiffany Ngyuyen for their patience and skill with a temperamental eye tracker. Melina Kunar provided helpful suggestions of improving the robustness of contextual cueing. Funding for this research was provided by the UCSD Temporal Dynamics of Learning Center (NSF award SBE-0542013, Garrison Cottrell, PI).

References

Armel, C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*, 396–403.

Azevedo, R., & Lajoie, S. P. (1998). The cognitive basis for the design of a mammography interpretation tutor. *International Journal of Artificial Intelligence in Education*, *9*, 32–44.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.

Dewhurst, R., & Crundall, D. (2008). Training eye movements: Can training people where to look hinder the processing of fixated objects? *Perception*, *37*, 1729–1744.

Donovan, T., Manning, D. J., Phillips, P. W., Highman, S., & Crawford, T. (2005). The effect of feedback on performance in a fracture detection task. *Progress in Biomedical Optics and Imaging (SPIE)*, *6*, 79–85.

Feldon, D. F. (2007). The implications of research on expertise for curriculum and pedagogy. *Educational Psychology Review*, *19*, 91–110.

Getty, D. M., Pickett, R. M., D'Orsi, C. J., & Swets, J. A. (1988). Enhanced interpretation of diagnostic images. *Investigative Radiology*, *23*, 240–252.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: guiding attention guides thought. *Psychological Science*, *14*, 462–466.

Guliato, D., Baventura, R. S., Maia, M. A., Rangayyan, R. M., Simedo, M. S., & Macedo, T. A. A. (2009). Indiam: An e-learning system for the interpretation of mammograms. *Journal of Digital Imaging*, *22*, 405–420.

Guthrie, E. R. (1952). *The psychology of learning (revised edition)*. Massachusetts: Harper Brothers.

Koning, B. B. de, Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, *20*, 111–122.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human Computer Studies*, *65*, 911–930.

Kunar, M. A., Flusberg, S. J., & Wolfe, J. M. (2006). Contextual cueing by global features. *Perception and Psychophysics*, *68*, 1204–1216.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied*, *16*, 251–262.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (p. 1-55). Hillsdale, NJ: Erlbaum.

Pashler, H., & Mozer, M. C. (2012). *When does fading enhance percpetual category learning?* (Submitted for publication)

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217.

Sutton, R. S., & Barto, A. G. (1988). *Reinforcement learning*. Cambridge, MA: MIT Press.

Thomas, L. E., & Lleras, A. (2009). Covert shifts of attention function as an implicit aid to insight. *Cognition*, *111*, 168–174.

Tomlinson, M. T., Howe, M., & Love, B. C. (2009). Seeing the world through an expert's eyes: Context-aware display as a training companion. In *Foundations of augmented cognition, neuroergonomics, and operational neuroscience (lnai 5638)* (pp. 668–677). Berlin: Springer.

Whitehill, J., & Movellan, J. (2012). Discriminately decreasing discriminability with learned image filters. In *Computer vision and pattern recognition* (pp. 2488–2495). Providence, RI: IEEE Xplore.

Wilder, M. H., Mozer, M. C., & Wickens, C. D. (2011). An integrative, experience-based theory of attentional control. *Journal of Vision*, *11*, 1-30.

Wright, R. D., & Richard, C. M. (1999). The effects of practice on location cueing and visual search. *Canadian Journal of Experimental Psychology*, *53*, 257–261.