

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Visual search guidance is best after a short delay

Joseph Schmidt, Gregory J. Zelinsky*

Stony Brook University, Stony Brook, NY, United States

ARTICLE INFO

Article history:

Received 5 November 2010
 Received in revised form 20 January 2011
 Available online 2 February 2011

Keywords:

Visual search guidance
 Consolidation
 Eye movements
 Visual working memory
 Target representation

ABSTRACT

Search displays are typically presented immediately after a target cue, but in the real-world, delays often exist between target designation and search. Experiments 1 and 2 asked how search guidance changes with delay. Targets were cued using a picture or text label, each for 3000 ms, followed by a delay up to 9000 ms before the search display. Search stimuli were realistic objects, and guidance was quantified using multiple eye movement measures. Text-based cues showed a non-significant trend towards greater guidance following any delay relative to a no-delay condition. However, guidance from a pictorial cue increased sharply 300–600 ms after preview offset. Experiment 3 replicated this guidance enhancement using shorter preview durations while equating the time from cue onset to search onset, demonstrating that the guidance benefit is linked to preview offset rather than a more complete encoding of the target. Experiment 4 showed that enhanced guidance persists even with a mask flashed at preview offset, suggesting an explanation other than visual priming. We interpret our findings as evidence for the rapid consolidation of target information into a guiding representation, which attains its maximum effectiveness shortly after preview offset.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The performance of common visuo-motor tasks, such as making a sandwich or assembling a piece of furniture, requires holding representations of specific utensils or tools in visual working memory (VWM) across short delays so as to fluidly coordinate interactions with these objects (e.g., Aivar, Hayhoe, Chizk, & Mruczek, 2005; Ballard, Hayhoe, & Pelz, 1995; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; see also Hayhoe & Ballard, 2005, for a review). Search is the perfect example of this. Just about any moderately complex visuo-motor task includes multiple search tasks as subordinate components. Most of these searches take place outside of our conscious awareness. When we place a screwdriver on the floor with the intent of reacquiring it a moment later, that is a search task. The properties of the screwdriver are held in VWM (Hollingworth & Luck, 2009; Woodman, Luck, & Schall, 2007), where they are used to efficiently guide search back to the target when the tool is again needed. This search-related function is supported by the finding of “look ahead” fixations (Hayhoe et al., 2003; Mennie, Hayhoe, & Sullivan, 2007; Pelz & Canosa, 2001). In preparation for a task, people often first look to the objects that will be required to perform the task. These fixations might very well mediate the temporary representation of these objects in VWM, with one function of these representations being to serve as target templates to

guide search to these objects as they are needed. This intertwined relationship between search and VWM means that target representations must often be held in memory over delays of perhaps several seconds. In this study we attempt to better understand the effect of these delays on search guidance.

Theories of visual search typically have not focused on the possibility that target representations may change over time; they instead assume an immediate and static representation of search targets, to the extent that their representation is discussed at all (e.g., Duncan & Humphreys, 1989; Treisman & Sato, 1990; Wolfe, 1994). This is especially true of the new breed of image-based search theories, which not only assume an unchanging target representation, but also that this representation consists of highly detailed visual information (Pomplun, 2006; Rao, Zelinsky, Hayhoe, & Ballard, 2002; Zelinsky, 2008). All of these theories posit that a target, typically specified using a pictorial preview, is compared to a search scene, with the outcome of this comparison operation then used to prioritize the acquisition of objects during search. The representation of the search scene is assumed to be perceptual, but the target preview (which would no longer be in view) must be represented in memory, presumably VWM (Hollingworth & Luck, 2009; Woodman et al., 2007). Given that VWM representations (or the accessibility of these representations) are known to change over time (e.g., Paivio & Bleasdale, 1974), changes to the target representation are therefore expected, which might affect how efficiently search is guided to the target. Depending on how these target representations change over time, existing search theories may be systematically

* Corresponding author. Address: Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, United States.

E-mail address: Gregory.Zelinsky@stonybrook.edu (G.J. Zelinsky).

overestimating or underestimating the availability of target information in the guidance process.

There are two obvious ways that a target representation in VWM might change over time. One possibility is that these representations might lose their fidelity over a delay. Representations in VWM are known to decay over the course of several seconds (e.g., Cornelissen & Greenlee, 2000; Eng, Chen, & Jiang, 2005; Paivio & Bleasdale, 1974), and the target representations used to guide search might suffer a similar fate. To the extent that this is true, search guidance would be expected to decrease with increasing delay between target designation and the actual search task. A second possibility is that VWM representations might become increasingly refined and elaborated over a delay. Just as information is believed to be consolidated in LTM over extended delays, particularly during sleep (e.g., Ellenbogen, Payne, & Stickgold, 2006; Payne, Stickgold, Swanberg, & Kensinger, 2008), a very rapid process of consolidation may occur in VWM, producing representations capable of mediating detection tasks after only a few hundred milliseconds have elapsed (e.g., Vogel & Luck, 2002; Vogel, Woodman, & Luck, 2006; Woodman & Vogel, 2008; see also Potter, 1976; Potter & Faulconer, 1975). To the extent that target representations are similarly consolidated in VWM, search guidance might be expected to improve with increasing delay between the target cue and the search display (cue–search delay).

Surprisingly few studies have investigated the relationship between cue–search delay and how efficiently search is guided to a target, with most studies opting for either an arbitrary delay period (e.g., one second; Zelinsky, 1999) or no delay whatsoever (e.g., Schmidt & Zelinsky, 2009). In the first study to explicitly address this relationship, Meyers and Rhoades (1978) examined search efficiency as a function of cue–search delay for stimulus onset asynchronies (SOAs) ranging up to two seconds. Their target cues were either pictures of objects or semantically-defined text labels, and search displays were realistic scenes. They found that search was most efficient for both pictorially-cued and textually-cued targets with an SOA of 500 ms, a finding that they interpreted as evidence for an optimal target encoding duration; it takes about 500 ms to effectively encode target relevant information from a cue and to prepare for search.

More recently, Wolfe and colleagues (Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004) also investigated the effect of cue–search delay for pictorially-cued and textually-cued targets. The primary goal of their study was to determine the time needed to load a target template for search, and towards this end they compared search efficiency when a target changed on every trial to a condition in which the target type was blocked. They also varied the SOA between target cue onset and search display onset, under the assumption that SOA might reveal the time needed to reconfigure the visual system when targets change from trial to trial. Two findings from this study are relevant to the current investigation. First, they found that search efficiency was consistently better with a pictorial target cue compared to a textual cue regardless of delay (see also Castelano, Pollatsek, & Cave, 2008; Meyers & Rhoades, 1978; Schmidt & Zelinsky, 2009; Vickery, King, & Jiang, 2005; and Yang & Zelinsky, 2009, for studies showing a general advantage of pictorial target cues over semantic target cues). Second, they found that the effect of delay depended on the type of target cue. Search efficiency with a semantic cue, although overall relatively poor, improved with longer SOAs. In contrast, a U-shaped relationship between search efficiency and SOA was found for pictorial cues; search efficiency first increased with SOAs up to about 200 ms, then decreased with longer delays. These patterns, which were independently replicated in subsequent work (Vickery et al., 2005), were interpreted as evidence for different delay-dependent processes acting

on pictorial and semantic cues. Given the converging trajectories of the semantic and pictorial delay effects, Wolfe and colleagues further speculated that the efficiency difference between cue types might disappear with an SOA of about 1600 ms, although this was never tested.

Despite their different stimuli and delay conditions, the findings from the Meyers and Rhoades (1978) and the Wolfe et al. (2004) studies were remarkably consistent. Both studies found a U-shaped relationship between search efficiency and SOA with pictorial cues (although the peak expressions of this efficiency occurred at slightly different delays), and the authors of these studies even reached similar conclusions. The target representation derived from a semantic cue was thought to become steadily elaborated over time, thereby explaining the improved search efficiency with increasing SOA. In contrast, the representations derived from pictorially previewed targets were thought to simply decay over time, producing a negative relationship between SOA and search efficiency. Importantly, no appeal was made to a feature consolidation process to explain the improvement in search efficiency observed after a short SOA. Rather, this U-shaped dip in the reaction time (RT) \times delay function was interpreted as evidence for an encoding limitation; with a brief target preview and a short SOA, there was simply insufficient time to encode enough of the target's features to efficiently guide search. The specific shape of the delay function is therefore believed to depend on the type of cue used to designate a target, semantic or pictorial, and ultimately on the recruitment of qualitatively different processes (e.g., encoding, decay, elaboration) at different times during the delay.

We had three goals for the present study. First, we hoped to explicitly test previous suggestions that cue–search delay effects are an artifact of target preview encoding limitations. Although incomplete target encoding is a reasonable explanation for the dip in the RT \times delay function previously found using pictorial cues and short SOAs, it is worth noting that both the Meyers and Rhoades study and the Wolfe et al. study presented target cues very briefly, thereby creating the very conditions that might produce an encoding limitation. We evaluate this possibility by presenting target cues for a full three seconds, which should allow for a more complete encoding of targets. If previous studies were correct in attributing this dip in the delay function to an early encoding limitation (leading to initially inefficient search), lifting this limitation should cause the dip to disappear. However, if the dip remains despite a long presentation of the target cue, that would suggest a process time-locked to cue offset rather than a limitation introduced at encoding. Second, we sought to better pinpoint the maximum expression of these effects. Although the Meyers and Rhoades (1978) and the Wolfe et al. (2004) studies collectively explored a wide range of cue–search delays, such comparisons are best conducted in the context of a single study that uses the same stimuli and methods. Relatedly, earlier work often repeated stimuli over trials (e.g., Vickery et al., 2005; Wolfe et al., 2004), thus introducing the potential for target memory from previous trials to affect the formation and maintenance of target representations over cue–search delays. We avoided this potential confound by never repeating stimuli in our study. Third, previous work quantified search efficiency purely in terms of manual RTs. While informative, this dependent measure left open the possibility that delay-related changes in search efficiency might reflect changing decision criteria rather than true search guidance; observers may have looked at distractors for longer or shorter durations as a function of delay, without differentially guiding their search to the target. By quantifying search efficiency in terms of eye movements, we can determine, unambiguously, whether delay affects the actual selection of search targets.

2. Experiment 1

In this experiment we seek to reinterpret the characteristic dip in the pictorial preview cue–search delay function in terms of two component processes. It might be the case that guidance improves shortly after cue offset due to the rapid assembly of target visual features into an efficient guiding representation by a process that is time-locked to preview offset. We will refer to this process as *consolidation*. Search guidance might also worsen over a delay due to the target's representation slowly degrading in VWM, a process that we will refer to as *decay*.¹ Given that these consolidation and decay processes need not have the same time course, the possibility exists that the dip in the RT \times delay function might result from a rapid process of consolidation followed by a slower process of decay. We will refer to this as the *consolidation–decay hypothesis*. Competing with this explanation is the suggestion that this dip is due to observers having an insufficient opportunity to encode the target preview's features with very short SOAs (Meyers & Rhoades, 1978; Wolfe et al., 2004), followed again by a gradual process of decay. We will refer to this possibility as the *encoding limitation plus decay hypothesis*.

As suggested in previous work, the type of process at work over a cue–search delay might also depend on the type of target cue. In the case of semantic cues, target representations must be self generated, presumably from object-specific features retrieved from LTM. Because such target representations could be refreshed from LTM at any time, we would not expect them to decay meaningfully over a delay. Moreover, assuming that only a small number of features from LTM are used to construct these representations, there would be no repository of high-fidelity visual information to consolidate, as assumed by the consolidation–decay hypothesis, and consequently no consolidation related guidance benefits. We therefore predict that the proportion of initial saccades directed to semantically-cued targets will not change with delay, or might increase only slightly with delay due to some minimal elaboration of the target's features (as Vickery et al., 2005 and Wolfe et al., 2004 observed for RTs). Pictorial target cues are likely to produce a very different pattern. The many visual details immediately available from a target preview create a greater opportunity for the expression of an encoding limitation, or the rapid consolidation of visual features in VWM, followed by the ultimate decay of this representation. If an encoding limitation was responsible for the relatively poor search efficiency found at very short cue–search SOAs, as previous studies have speculated, then the long preview durations used in the present study should flatten the delay function; search should be maximally efficient with very short (or zero) cue–search delays, and steadily decline with increasing delay. With regard to eye movement evidence for search guidance, the encoding limitation plus decay hypothesis therefore predicts a high proportion of initial saccades to previewed targets immediately after preview offset, followed by a monotonic decline in search guidance as the target representation gradually decays. Importantly, we should find no peak in the search guidance \times delay function. However, if the previously reported dip in the delay function reflected an actual improvement in search guidance after a short cue–search delay, as predicted by the consolidation–decay hypothesis, it should remain despite the longer opportunity to encode the target preview. In this case, we would expect an inverted U-shaped pattern of initial saccades to the target in the delay function, consistent with the pattern of manual RTs reported previously. Of course, this pattern could no longer be realistically attributed to an encoding limitation.

¹ Whereas we use the term “decay” to describe the worsening of guidance over a delay, we do this as a matter of convenience and do not distinguish in this study between decay-based and interference-based explanations of forgetting from VWM.

2.1. Methods

2.1.1. Participants

Thirty-two undergraduate students from Stony Brook University participated for course credit. All had normal or corrected-to-normal vision, and were native English speakers, by self-report.

2.1.2. Stimuli and apparatus

Stimuli consisted of a target cue and a 5-item search display (Fig. 1). Target cues were either a pictorial preview of the target or a text label describing the target, both presented at central fixation on a white background. The text labels were centered in a 3° white box and were written in 18-point black Tahoma regular font. Search displays were presented on a 20° \times 20° white background, and consisted of four distractors and one target (all Hemera Photo Objects). The objects were arranged in a circle (9° radius from central fixation), and were positioned by first placing the target at a random point on the circle's circumference, then placing each distractor at 72° increments along the circle relative to the target's location. Each object subtended approximately 2°.

A small (.14°) + or \times character was inserted next to the target in the search array using Adobe Photoshop CS, black Tahoma regular (crisp) 7-point font. Over trials, half of the targets appeared with a + and the other half with a \times . These characters were manually placed as close to the target as possible, either touching or within two pixels. Positioning around the target was random, with the constraint that they were legible and easy to segment, while still being difficult to discriminate unless directly fixated.

Eye position was sampled at 500 Hz, using an EyeLink II eye-tracker (SR Research) with default saccade detection settings. Head position and viewing distance were fixed at 72 cm using a chinrest. Judgments were made by pressing the left and right index finger triggers of a game pad controller; trials were initiated with a button operated by the right thumb.

2.1.3. Design and procedure

There were two target cue conditions (between-subjects variable). The *Pictorial* condition showed a pictorial preview of the target. The *Semantic* condition showed a text label describing the target, which emphasized object shape, color, and details of the object's category (see Fig. 1). Both types of cues were displayed for 3000 ms to maximize cue encoding and elaboration before the start of the delay period. There were 68 experimental trials per cue type, and search displays were identical across cue conditions. There were also three inter-stimulus interval (ISI) conditions (600 ms, 3000 ms, and 9000 ms) and one no-ISI (0 ms) condition. These were blocked (within-subject variable) and counterbalanced across observers and trials.

Targets and distractors never repeated, and each trial used a different target category. This stimulus set was designed to prevent target categorical overlap across trials so as to minimize the potential for priming, interference, and bias effects which could impact search performance in the semantic text label condition. Had we not taken this precaution, using “rocking chair” as a target on an earlier trial might bias observers cued with “chair” on a later trial to search for a rocking chair, given their previous exposure to this target category. Target categories were also prevented from overlapping with distractors, and categorical overlap was even minimized within the distractors by using 47 different distractor categories.

The experiment began with a calibration routine used to map eye position to screen coordinates. A calibration was not accepted until the average error was less than .4° and the maximum error was less than .9°. Following calibration were eight practice trials, used to familiarize observers with the task and stimuli, as well as assess whether people could differentiate between the + and \times

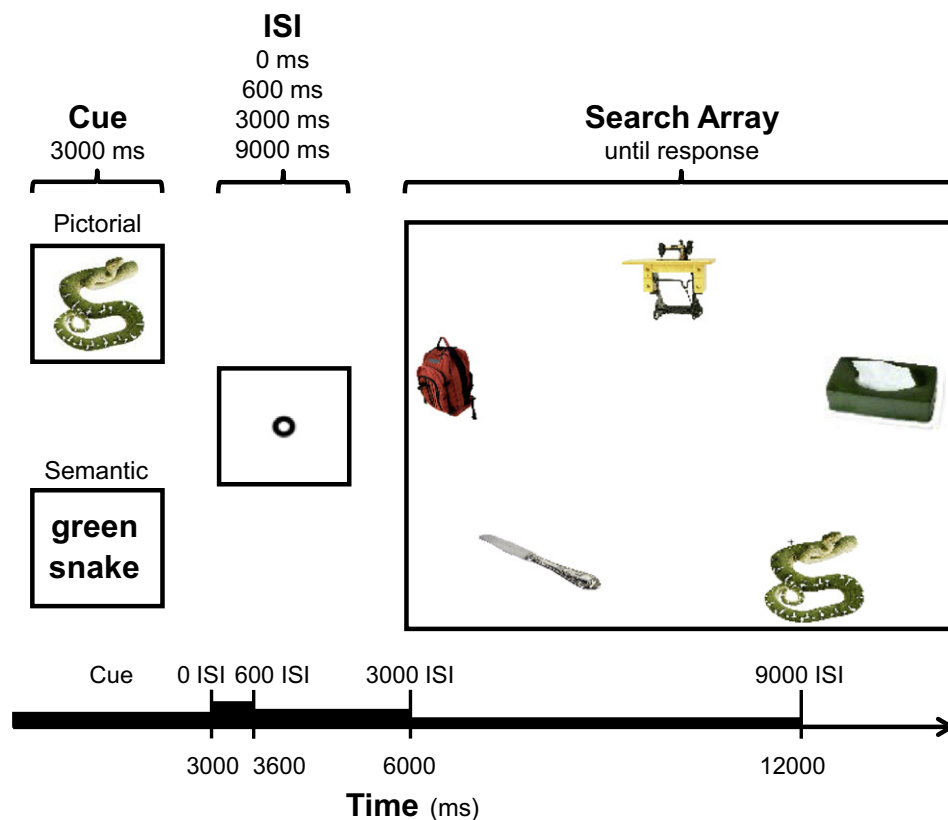


Fig. 1. Cuing and inter-stimulus interval (ISI) conditions used in Experiment 1 in the context of a representative search trial. Cues were presented for 3000 ms followed immediately by the search display (no-delay condition; 0 ms ISI) or by ISIs of 600 ms, 3000 ms, or 9000 ms before search array onset. Corresponding SOAs between cue onset and search array onset were therefore 3000 ms, 3600 ms, 6000 ms and 12000 ms, however we will only refer to these conditions by their respective ISI durations in this and the following experiments.

characters. To allow people to become accustomed to the various delays, the same eight practice trials were repeated at the corresponding ISIs before each block of trials.

Trials began with observers fixating a central point and pressing a button on the game pad. In addition to starting the trial, this served as a “drift correction” for the eye-tracker to account for any movement since calibration. The fixation point was then replaced by the target cue, which was displayed for 3000 ms to ensure that observers had adequate time to encode the target. For no-ISI trials, the search display appeared immediately after the offset of the target cue. On all ISI trials, the cue was replaced by a fixation point that remained until the search display. This fixation point was displayed for either 600 ms, 3000 ms or 9000 ms, depending on the ISI condition. An auditory warning tone sounded 600 ms before search display onset on all ISI trials. For no-ISI trials, the tone sounded before the preview rather than 600 ms before search array onset. This was done so as to prevent the tone from sounding during the target cue, which potentially could have disrupted encoding. The search display was then presented and remained until the discrimination judgment. Given our focus on target guidance following a delay, target absent trials would be relatively uninformative. For this reason a target was present on every trial, and observers were instructed to first find the target, as quickly and as accurately as possible, then to indicate the presence of a + or × by pressing the left or right triggers, respectively. This discrimination task has been successfully used in previous work that also focused on target guidance (Chen & Zelinsky, 2006; Schmidt & Zelinsky, 2009). The entire experiment lasted approximately 40 min.

2.2. Results

Errors were uncommon in this task, averaging less than 4% in all conditions. This indicates that observers were generally able to correctly identify the + and × characters. Error trials were excluded from all subsequent analyses.

To gauge search guidance across conditions, we analyzed the direction of the initial saccade following search display onset, the initial saccade latency, and the number of eye movements made before the first fixation on the target. The direction of the initial saccade is one of the most stringent measures of guidance because observers would have only a couple hundred milliseconds to analyze the search display before making this eye movement (Chen & Zelinsky, 2006). Across trials, a greater percentage of initial saccades directed to the target would indicate fast acting search guidance that is likely driven by lower-level perceptual processes. The initial saccade latency is a measure of the time needed to analyze the search display so as to produce the obtained level of initial saccade guidance. It also provides a reaction time measure for the initial saccade, which could indicate a speed–accuracy tradeoff with guidance. The number of eye movements to the target was analyzed so as to evaluate guidance that was not available in the first few hundred milliseconds of the search process, but rather developed upon closer inspection of the search display. However, this measure may also reflect a greater contribution of non-guidance related processes, such as those used in object identification and higher-level decision making.

The direction of the initial saccade was calculated as follows. First, the imaginary circle upon which the search objects appeared

was divided into five equal-sized 72° slices, one for each object. We then determined the vector of the first 2° (or greater) saccade after search display onset, and projected it onto the imaginary circle to obtain the point of intersection. If the projected saccade intersected the target section, it was counted as an eye movement toward the target. By chance, 1/5 or 20% of these initial saccades should be directed to the target; a significantly greater preference to saccade toward the target would indicate search guidance.

The encoding limitation plus decay hypothesis predicts that guidance from a pictorial cue should be strong with little or no delay, but then should decrease monotonically with increasing delay due to visual features fading from VWM. The consolidation–decay hypothesis predicts that guidance from a pictorial cue should start off relatively low, but then increase with delay as target features are abstracted into a more durable and representative code. This initial guidance benefit should then decrease with longer delays, as these visual details decay from VWM. For the semantic cue condition we predicted that guidance should change relatively little with delay, or perhaps even increase slightly over time as additional target features are self generated and added to the guiding target template. We also expected that guidance should be better with a pictorial cue than a semantic cue at all but the longest delay.

Fig. 2 shows the percentage of initial saccades directed to the target plotted as a function of cue condition and ISI, the delay between the offset of the 3000 ms cue and the onset of the search array. As expected, the pictorial preview condition resulted in significantly more initial saccades directed to the target than the semantic text label condition, $F(1, 30) = 5.75, p = .02$. This difference confirms previous work, indicating the general superiority of a pictorial cue in guiding search within the first few hundred milliseconds of a search task. However, it is also clear that the effect of cue type interacted with ISI, $F(3, 90) = 3.02, p = .04$. Consistent with our expectation, a separate analysis of the semantic cue data showed only a non-significant trend toward better guidance with longer ISIs, $F(3, 45) = 1.43, p = .25$. In contrast, a separate analysis of the pictorial preview data revealed a significant effect of ISI, $F(3, 45) = 2.99, p = .04$; the percentage of initial saccades directed to the target was greater after a 600 ms delay compared to after a 9000 ms delay ($p = .01$; all post-hoc tests presented throughout this paper used Least Significant Difference (LSD) pairwise comparisons). The contrast between the 600 ms and 0 ms ISI conditions also approached significance ($p = .07$). Search guidance from a pictorial target cue did not simply decrease monotonically with

increasing cue–search delay, but rather first increased, then decreased, peaking at 600 ms in this experiment.

To more closely investigate this guidance benefit of pictorial cuing over semantic cuing we conducted independent samples *t*-tests on the initial saccade direction data at each level of ISI. Significant differences across cue type in this measure would indicate an advantage of the pictorial cue very early in the search process. We found that significantly more initial saccades were directed to the target after a 600 ms ISI in the pictorial condition, $t(31) = 3.74, p = .001$, and trends toward similar differences existed at the 0 ms and 3000 ms ISIs as well, although these were not reliable. However, there was not even the suggestion of a guidance difference between pictorial and semantic conditions after a 9000 ms ISI ($p = .95$). Pictorial search guidance, as measured by initial saccade direction, started off only slightly better than semantic guidance, rapidly increased to produce a significant difference after a 600 ms ISI, then declined until there was no difference between the two cuing conditions after a 9000 ms ISI. This data pattern is what you would expect from a semantically-driven guidance process that changes relatively little with ISI, and a pictorially-driven guidance process that peaks after a short ISI, but drops off at longer delays.

It is possible that longer initial saccade latencies might explain the more accurately directed initial saccades observed in the pictorial 600 ms ISI condition; if these saccades were delayed, more opportunity would exist to accumulate information about the search scene that might be used to better guide the saccade to the target. However, post-hoc analyses following ANOVA revealed that initial saccade latencies after a 600 ms ISI were not systematically longer than those in the other pictorial cue conditions (all $p > .05$, Table 1). We also found that initial saccade latencies did not differ significantly between cue types, $F(1, 30) = .43, p = .52$. Latencies did differ across our pictorial ISI conditions, $F(3, 45) = 7.55, p < .001$, but this difference was carried by longer initial saccade latencies in the 0 ms ISI condition compared to the others (all $p \leq .05$). A speed–accuracy tradeoff can therefore not explain the guidance benefit found in the pictorial condition after a 600 ms delay.

Is the above-described guidance benefit specific to the initial saccade made during search? If so, its overall importance to the search process might be questioned. To address this, we analyzed the number of eye movements to fixate the search target, a measure of guidance that better reflects the entirety of the search process. These data appear in Table 2. Complimenting the initial saccade direction results, this measure of guidance also revealed a clear advantage for the pictorial cue, $F(1, 30) = 9.83, p < .001$, as well as a main effect of cue–search delay, $F(3, 90) = 4.20, p = .009$, and a marginally significant cue \times delay interaction, $F(3, 90) = 2.52, p = .06$. Analyzing the effect of ISI separately for each cue condition we found that ISI did affect the number of eye movements to the target with a pictorial cue, $F(3, 45) = 7.26, p < .001$. After a 600 ms ISI observers made significantly fewer saccades to reach the target compared to the 0 ms and 9000 ms ISI conditions (both $p < .02$). Significantly fewer eye movements were also required to fixate the target after a 3000 ms ISI compared to

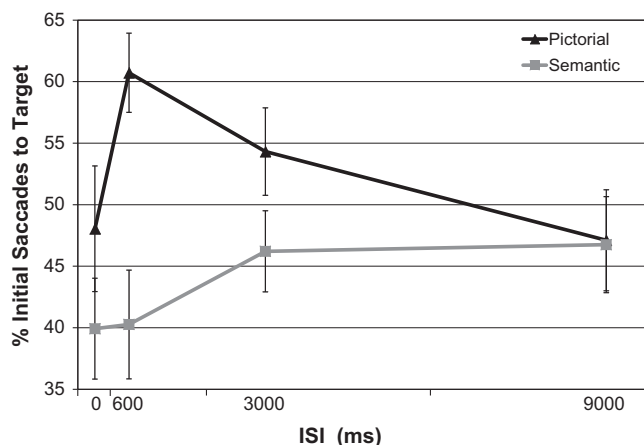


Fig. 2. Percentage of initial saccades directed to the target in Experiment 1 (correct trials only), as a function of cue condition and delay between the offset of the cue and the onset of the search array. Error bars indicate one standard error of the mean (SEM).

Table 1
Initial saccade latencies (ms) for correct trials in Experiment 1.

ISI (ms)	Pictorial	Semantic
0	201 (4.6)	199.1 (6.1)
600	186 (8.5)	190.8 (8.9)
3000	179 (5.0)	194.9 (12.4)
9000	173 (5.8)	179.5 (6.6)

Note: Values in parentheses indicate standard error of the mean (SEM).

Table 2
Number of eye movements to the target for correct trials in Experiment 1.

ISI (ms)	Pictorial	Semantic
0	2.2 (0.11)	2.9 (0.23)
600	1.9 (0.08)	2.5 (0.14)
3000	2.0 (0.09)	2.6 (0.23)
9000	2.4 (0.13)	2.6 (0.14)

Note: A target was considered fixated if gaze landed either on the object or within 1.1° of its outer contour. Values in parentheses indicate standard error of the mean (SEM).

the 9000 ms ISI condition ($p = .009$). The number of eye movements to the target did not vary reliably with ISI in the semantic cue condition, $F(3, 45) = 1.49$, $p = .23$, although there was again a non-significant trend towards improved guidance at longer delays. Comparing the pictorial and semantic conditions, we also found that the pictorial cue produced additional guidance over the semantic cue at all but the longest delay, $t(30) \geq 2.59$, $p \leq .02$; after a 9000 ms ISI the number of saccades to the target did not significantly differ, $t(30) = .90$, $p = .38$. Together, these analyses are perfectly consistent with the initial saccade direction results in suggesting that both ISI and cue type affect search guidance. Fewer eye movements were needed to locate the target with a pictorial cue, and with a pictorial cue guidance was best after a 600 ms delay.²

2.3. Discussion

To summarize, analyses of the initial saccade directions and the number of eye movements to the target both revealed the same general pattern; a peak in pictorial search guidance after a 600 ms delay, followed by a decline in guidance with longer delays. The semantic cue condition showed no such peak, exhibiting only a non-significant trend towards better guidance with longer, non-zero ISIs. Guidance with a pictorial cue was also generally superior to guidance with a semantic cue. However, this superiority did not hold for the longest delay; as the level of pictorial guidance decreased following its 600 ms ISI peak, this difference between the pictorial and semantic cue conditions eventually disappeared.

We believe that the decline in the pictorial condition guidance following longer (>600 ms) delays is likely due to visual features fading from VWM. More specifically, we speculate that the decay of target-related visual details begins approximately 600 ms after preview offset, and continues until the target information used to guide search degrades into a representation qualitatively similar to the one formed from an elaborated semantically-defined cue. In contrast, target representations formed from semantic cues would not be expected to similarly degrade over comparable delays. Because the creation of a guiding representation from these cues likely involves the assembly of target features from one's semantic knowledge of a target class (Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009), there would be no veridical visual details to decay from VWM. Indeed, target cues constructed from LTM might become more effective in guiding search over time, due to the increased opportunity to elaborate upon these cues. However, this guidance benefit would likely be small and limited by one's ability to add useful visual details to the target template (Schmidt & Zelinsky, 2009), a pattern consistent with our observation of only a non-significant trend towards better guidance with increased delay under semantic cuing conditions. Prior reports of significantly

improved search efficiency following a delay and a semantic cue likely resulted from too little time to read and elaborate upon the cue during its presentation (Meyers & Rhoades, 1978; Vickery et al., 2005; Wolfe et al., 2004); the much longer cue duration used in the present study would have allowed for these processes to largely complete before the delay manipulation.

In contrast to the slightly improved guidance over a delay in the semantic cue condition, guidance in the pictorial cue condition was substantially improved following a short delay. Previous work attributed this delay-dependent improvement in search efficiency to the longer opportunity to more completely encode the target (Meyers & Rhoades, 1978) or to set top-down weights to guide search (Wolfe et al., 2004). At short delays these processes would be truncated by the search display, resulting in poor search efficiency. However, with longer delays these processes would be able to continue, resulting in a more complete or optimized guiding representation. This explanation in terms of an encoding limitation seems unlikely given the duration of the target previews used in the present experiment. Meyers and Rhoades used preview durations of 200 ms, and Wolfe and colleagues used preview durations as short as 50 ms; our preview durations were 3000 ms, 15–60 times longer than those used in the previous work. These dramatically longer preview durations would be expected to alleviate any limitation of target encoding. Guidance should therefore be strongest immediately following preview offset, before any delay-related decay of the target representation, and decrease monotonically with increasing preview–search delay. This prediction of the encoding limitation plus decay hypothesis was clearly inconsistent with our data; guidance was not best immediately following the preview, but rather peaked shortly thereafter. Despite our dramatically longer cue presentation times, we still found that a short delay following preview offset momentarily improved search efficiency.

We believe that the consolidation–decay hypothesis offers a more satisfying explanation for our data. According to this hypothesis, the improvement in search efficiency shortly after preview offset, defined in our current data by a difference between the no ISI and 600 ms ISI conditions, reflects an actual period of momentarily enhanced guidance (and not an encoding artifact) due to a rapid consolidation of target features in VWM that is time-locked to preview offset. We will refer to this rapid boost in target guidance as the *guidance enhancement effect* (GEE). Following this initial boost, the consolidation–decay hypothesis asserts that the target representation responsible for guidance should begin decaying from VWM. This decay process is captured in our data by differences between the 600 ms and 9000 ms ISI conditions. Collectively, these consolidation and decay processes produce the peak that characterizes the guidance \times delay function. In the following experiment we seek to better localize this peak, thereby better pinpointing the maximal expression of the GEE.

3. Experiment 2

In Experiment 1 we determined that pictorial search guidance is significantly improved after a 600 ms delay, but that this improvement largely disappeared after a 3000 ms delay. This is a fairly broad range that may fail to pinpoint the GEE's strongest expression. For example, it may be the case that the GEE peaks significantly earlier or later than 600 ms, or that this effect should not be characterized as a peak at all; it may be that the GEE is a plateau of strong guidance that is sustained for a second or more. In Experiment 2 we explored different delay conditions so as to better determine how long after preview offset the enhancement begins, and how long it lasts.

² Similar results were obtained in an analysis of the time taken to fixate the target, but this analysis was omitted due to its redundancy with the other eye movement analyses.

3.1. Methods

3.1.1. Participants

Forty undergraduate students from Stony Brook University participated for course credit. All had normal or corrected-to-normal vision, by self-report, and none had participated in Experiment 1.

3.1.2. Design and procedure

Stimuli, apparatus, and procedure were identical to the descriptions provided in Experiment 1, with the following exceptions. To more fully explore a range of delays, we divided observers into two 20-participant groups. For one group, the ISIs between target preview offset and search display onset were 0 ms (no ISI), 50 ms, 400 ms, and 800 ms; for the other group these ISIs were 0 ms, 100 ms, 200 ms, and 300 ms. Only pictorial target previews were used in this experiment. We also changed when the auditory warning tone sounded. In Experiment 1 the tone sounded before the target preview on no-ISI trials, but sounded 600 ms before the onset of the search display on ISI trials. It is possible that this difference between the ISI and no-ISI conditions may have contributed to the GEE by creating different levels of arousal, attentional readiness, or vigilance. To remove this potential confound in Experiment 2 the auditory warning tone sounded 600 ms before the search display on all trials. If differences in arousal-related factors between the no-ISI and ISI conditions were contributing to the GEE's expression in Experiment 1, we would expect a substantially smaller GEE in Experiment 2.

3.2. Results and Discussion

In the pictorial condition from Experiment 1 we found that the prevalence of initial saccades to the target peaked after a short preview–search delay. This increase in initial saccades to the target constitutes a very strict measure of search guidance, one that is most likely to minimize the contribution of post-perceptual factors. Because of this, all remaining analyses in this paper will characterize guidance enhancement exclusively in terms of this single dependent measure.

Fig. 3 plots the percentage of initial saccades directed to the target by delay condition. The triangle markers indicate data from the Group 1 observers; the circle markers indicate data from Group 2. Separate repeated-measures ANOVAs conducted on these groups each showed an effect of ISI on search guidance, $F(3, 57) \geq 2.77$, $p \leq .05$. Through post-hoc tests we determined that a 300 ms ISI

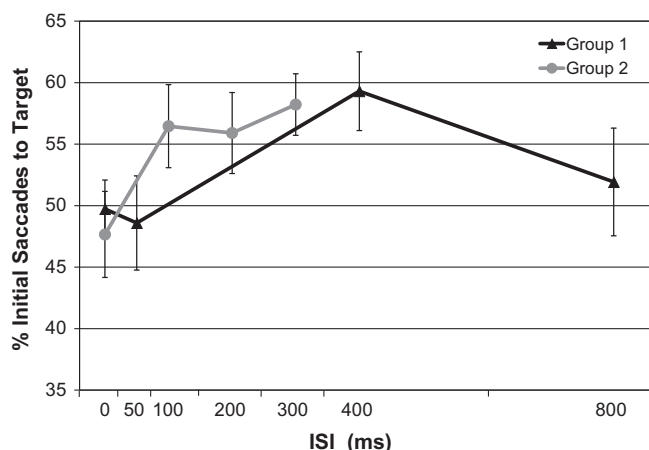


Fig. 3. Percentage of initial saccades directed to the target as a function of delay condition for correct trials in Experiment 2. Error bars indicate one standard error of the mean (SEM).

produced significantly more initial saccades to the target when compared to a 0 ms ISI ($p = .01$), and that a 400 ms ISI produced significantly more initial saccades to the target when compared to 0 ms, 50 ms, and 800 ms ISIs (all $p \leq .03$).³ When combined with the results from Experiment 1, these findings indicate that the GEE is maximally expressed over delays ranging from 300 to 600 ms. Guidance steadily increases over the first 300 ms after preview offset, but remains relatively stable between the 300 ms, 400 ms, and 600 ms delay conditions. Guidance then declines with longer delays, which we can now pinpoint as occurring as early as 800 ms following preview offset. We interpret these findings as evidence for the rapid consolidation of visual information into a guiding target representation, with this process reaching completion in approximately 300 ms following target preview offset. However, this representation is short-lived. The boost in guidance made possible by this representation is only available for about another 300 ms, after which it starts to degrade and guidance returns to its pre-enhancement levels.

4. Experiment 3

One of the most surprising properties of the GEE is that it peaked shortly after the target preview disappeared; guidance was not best immediately following preview offset. This finding raises the intriguing possibility that the target representations used to guide search are constructed relative to preview offset, not onset. Recall that the Wolfe et al. (2004) study and the Meyers and Rhoades (1978) study assumed that the construction of the guiding representation began at preview onset, but that the preview was presented for too short a time to bring this construction to completion. For this reason, processing was thought to continue following preview offset, resulting in the boost to overt search guidance that we termed the GEE. Although the dramatically longer target preview durations used in Experiments 1 and 2 make this encoding limitation explanation unlikely, these experiments did not explicitly rule out this possibility. Our interpretation of the GEE is quite different. We hypothesize that the target representation in our task is constructed relative to preview offset, as part of a process of feature consolidation in VWM. According to this hypothesis, manipulating preview duration should have relatively little effect on search guidance; the GEE should only be expressed with the insertion of a delay following preview offset.

Experiment 3 used two conditions to tease apart onset coding from offset coding, one having a 1400 ms preview duration and a 0 ms ISI (1400 + 0 condition), and the other having a 1000 ms preview duration and a 400 ms ISI (1000 + 400 condition). Note that the overall time between target preview onset and search display onset is equated in these conditions, with the only difference between them being that one has an ISI (and hence a shorter preview duration) than the other. To the extent that the onset coding hypothesis is correct, we would expect to find no difference in guidance between these conditions, as the delay relative to preview onset is held constant. Indeed, if guidance is affected at all, it might decrease in the 1000 + 400 condition due to the shorter preview duration providing less time to encode the target. However, to the extent that the offset coding hypothesis is correct, we would expect to find improved guidance in the 1000 + 400 condition, due to the 400 ms ISI following preview offset providing the opportunity for target feature consolidation in VWM. We also added a third condition, one having a 1000 ms preview duration

³ A qualitatively similar pattern of results was obtained in the number of fixations to the target and time to fixate the target, as in the case of Experiment 1. In pilot work we also found a GEE at short delays using an interleaved design, suggesting that the blocked design used throughout this study was not responsible for the GEE's expression.

and no ISI (1000 + 0 condition), so as to independently test for an encoding limitation using these shorter preview durations (recall that 3000 ms durations were used in Experiments 1 and 2). To the extent that the GEE is not an artifact of an encoding limitation (at least within the range of preview durations explored in this study), we would expect to find no guidance difference between the 1400 + 0 and 1000 + 0 conditions. In Experiment 3 we test these hypotheses.

4.1. Methods

4.1.1. Participants

Fifteen undergraduate students from Stony Brook University participated for course credit. All had normal or corrected-to-normal vision, by self-report, and none had participated in Experiments 1 or 2.

4.1.2. Design and procedure

Stimuli, apparatus and procedure were identical to Experiment 2, except for the use of the three above-described within-subjects conditions: A 1400 ms target preview with a 0 ms cue–search ISI (1400 + 0 condition), a 1000 ms target preview with a 400 ms cue–search ISI (1000 + 400 condition), and a 1000 ms target preview with a 0 ms cue–search ISI (1000 + 0 condition).

4.2. Results and discussion

As in Experiments 1 and 2, we quantified the GEE in terms of a difference in the direction of initial search saccades between the ISI and no-ISI conditions, in this case the 1000 + 400 condition and the 1000 + 0 condition. Consistent with the previous experiments, search after an ISI (1000 + 400) produced significantly more initial saccades to the target compared to search without an ISI (1000 + 0), $t(14) = 2.18$, $p = .05$ (Fig. 4). This finding indicates that the expression of the GEE does not require the long three second preview durations used in Experiments 1 and 2. We also compared the 1000 + 0 condition to the 1400 + 0 condition, and found essentially identical levels of guidance, $t(14) = 0.26$, $p = .80$. This finding suggests that the target was encoded equally well after 1000 ms and 1400 ms previews, and provides converging evidence against the possibility that the GEE was caused by a target encoding

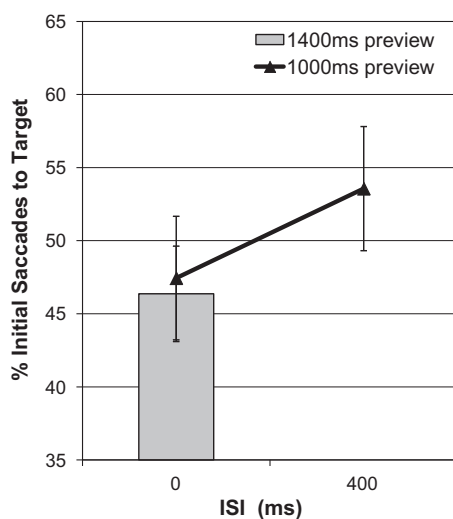


Fig. 4. Percentage of initial saccades directed to the target for correct trials in Experiment 3 as a function of delay and preview duration condition. Error bars indicate one standard error of the mean (SEM).

limitation. This is particularly true with respect to the data from Experiments 1 and 2; if no evidence for an encoding limitation was found with a one second target preview, one could not have existed with the three second previews used previously. With respect to the critical test between the onset encoding hypothesis and the offset encoding hypothesis, we found significantly improved guidance in the 1000 + 400 condition compared to the 1400 + 0 condition, $t(14) = 2.50$, $p = .03$. This difference strongly suggests that the delay following the offset of the target preview is responsible for the enhanced guidance, and not the delay relative to the preview's onset. Although features of the target are undoubtedly encoded at its onset, our data suggest that an analogous process occurs at target offset, with these offset features used to create a momentarily improved target representation resulting in a short-lived boost in search guidance.

5. Experiment 4

Experiment 3 demonstrated that the GEE is time-locked to target preview offset, leaving us with the working hypothesis that target disappearance elicits a process of rapid feature consolidation in VWM that results in a momentary boost in guidance to the target. If this hypothesis is correct, we might expect some effect of a visual mask inserted immediately at target preview offset, before this consolidation is believed to occur. One possibility is that such a mask might eliminate the GEE by selectively removing the information that is used by the consolidation process. This would be expected if the GEE was a form of delayed visual priming, one that takes several hundred milliseconds to fully exert itself (e.g., Wolfe et al., 2004). Another possibility is that a mask might degrade the target's features in VWM more broadly, but not completely erase this information. This should result in an equal reduction in guidance under ISI and no-ISI conditions, leaving the magnitude of the GEE relatively unchanged; search guidance after a mask would simply be offset relative to search guidance without a mask. In this experiment we tease apart these possibilities, thereby potentially revealing the mechanism for enhanced search guidance in our task.

5.1. Methods

5.1.1. Participants

Twenty undergraduate students from Stony Brook University participated for course credit. All had normal or corrected-to-normal vision, by self-report, and none had participated in the previous experiments.

5.1.2. Design and Procedure

Stimuli, apparatus and procedure were identical to Experiment 3, with the following exceptions. Preview durations were restricted to 1000 ms. Trials were evenly divided into mask and no-mask conditions, with each condition also divided into ISI and no-ISI trials. No-mask trials were identical to the 1000 ms preview duration trials from Experiment 3, except that the ISI was extended to 600 ms (from 400 ms). On mask trials, a 3° circular colored noise mask was flashed for 200 ms at preview offset. This means that no-ISI mask trials were 200 ms longer than no-ISI no-mask trials, due to the insertion of the 200 ms mask. To equate the time from preview offset to search display onset between the mask ISI and no-mask ISI trials, the delay on mask trials was shortened to 400 ms. Thus, from preview offset, a 200 ms mask followed by a 400 ms ISI produced a 600 ms total delay, equivalent to the preview–search delay used in the no-mask trials. Mask and ISI conditions were blocked and counterbalanced across observers and trials.

5.2. Results and discussion

Fig. 5 shows the percentage of initial saccades directed to the target with, and without, a mask and a delay. With regard to our hypotheses, neither is strongly supported. It is clear that inserting a mask after target preview offset did not eliminate the GEE, as indicated by a significant main effect of ISI, $F(1, 19) = 7.51$, $p = .01$. It is also clear that adding a mask did not simply offset the expression of the GEE, as indicated by the absence of a significant main effect of mask, $F(1, 19) = 2.75$, $p = .11$. However, although we failed to find a reliable mask \times ISI interaction, $F(1, 19) = 1.67$, $p = .21$, there is a suggestion in the data that a mask might selectively reduce search guidance only in the absence of a delay. To more thoroughly explore this relationship, we compared the masking conditions using separate paired-samples t -tests conducted on the no-delay and delay data. With no-delay, adding a mask resulted in a marginally significant reduction in search guidance, $t(19) = 2.05$, $p = .055$; with a delay, there was no evidence for a difference between the masking conditions, $t(19) = 0.81$, $p = .43$.

To the extent that a masking effect is specific to the no-ISI condition, how might we explain this highly counter-intuitive relationship between masking and cue-search delay? One plausible interpretation is that the mask injected noise into the target's representation (consistent with the above mentioned broad degradation hypothesis), but that this noise was adaptively removed as part of the consolidation process occurring over the delay. Without a delay, the opportunity to filter out noise from the mask would not exist, resulting in a degraded guidance signal and the suggestion of a difference between the masking conditions. However, with a delay this noise could be effectively removed, so much so that the level of guidance did not differ from what was observed in the no-mask condition following a delay. Although this interpretation is clearly speculative, what is certain from our data is that the addition of a mask following preview offset did *not* eliminate the GEE. Indeed, the masking cost to guidance found in the no-delay condition, combined with the near complete recovery of guidance found in the delay condition, resulted in the mask actually *increasing* the size of the GEE. This pattern would not be expected if the target preview was simply priming the target in the search display. Rather, these data are consistent with a process of feature consolidation time-locked to target preview offset. As part of this consolidation process, target visual information is analyzed

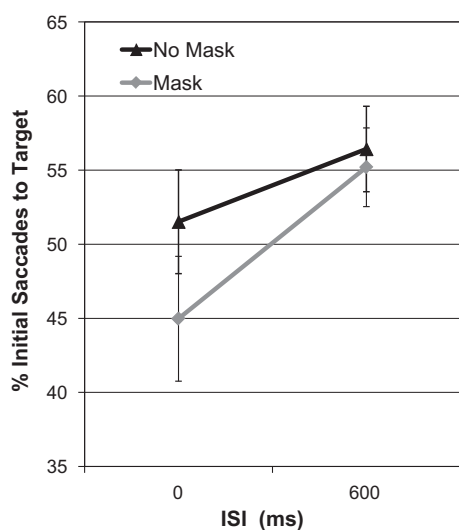


Fig. 5. Percentage of initial saccades directed to the target as a function of delay and masking condition for correct trials in Experiment 4. Error bars indicate one standard error of the mean (SEM).

at preview offset for the purpose of selecting (or assembling) the features needed to maximize search guidance.

6. General discussion

In this paper we explored how visual search guidance changes with the duration of a delay between the offset of the target cue and the onset of the search display. As part of this effort we showed that pictorially-cued targets are generally superior to semantically-cued targets (see also, Castelano et al., 2008; Meyers & Rhoades, 1978; Schmidt & Zelinsky, 2009; Vickery et al., 2005; Wolfe et al., 2004; Yang & Zelinsky, 2009), that semantically-cued targets produce a slight overall improvement in search efficiency with delay (see also, Vickery et al., 2005; Wolfe et al., 2004), and that the practice of repeating search items (e.g., Wolfe et al., 2004) seems not to have been critical in producing any of the previously reported delay effects, as we obtained qualitatively similar results without repeating stimuli.

Our adoption of eye movement dependent measures also enabled us to better specify the effects of cue-search delay on search. By analyzing the proportion of initial saccades made to the target, we determined that the boost in search efficiency following a short delay using a picture target cue reflects actual search guidance, and that this *guidance enhancement effect* (GEE) appears very early in the search process. These relationships were unclear from previous studies due to their exclusive use of manual dependent measures. Through a parametric exploration of cue-search delays we were also able to show that the GEE is maximally expressed after a delay of 300–600 ms, with this momentary boost in pictorial guidance disappearing completely after a 9000 ms delay. Interestingly, although guidance after a long delay was significantly lower than that after a short delay, it never dropped below what we observed at no delay. This pattern supports the characterization of the GEE as a momentary boost in guidance that fades with time, rather than a disruption in the guidance process when the search display immediately follows the target preview; it would be quite coincidental for a disruption at no delay to perfectly mimic guidance after a long delay.

Finally, perhaps the most significant new insight arising from this study is that the GEE is time-locked to preview offset. Previous work had attributed this boost in search efficiency to continued encoding of the target during the delay (Meyers & Rhoades, 1978; Wolfe et al., 2004), suggesting a process time-locked to preview onset, not offset. Because of the very short target previews used in these studies, this explanation, which appeals to an encoding limitation, was plausible. However, we manipulated preview duration and found that the expression of the GEE does not depend on the duration of the target cue. This result, combined with our repeated observation of the GEE using relatively long target previews, suggests instead a guidance process that peaks shortly after preview offset. Collectively, these findings are better explained by what we are calling the consolidation-decay hypothesis. Upon preview offset, the target features are rapidly consolidated into a VWM representation capable of mediating highly efficient search guidance. This consolidation process is followed by a slower process of decay from VWM, thereby producing a decline in guidance with increasing cue-search delay. This decay continues until the pictorial target representation becomes no better at guiding search than an elaborated target representation self-generated from a semantic cue. Determining whether this post-decay pictorial representation is informationally equivalent to an elaborated semantic representation will be a question for future work.

Why might search guidance be time-locked to preview offset? In one sense this seems wasteful, not taking full advantage of the availability of the target cue. However, when one considers the

suggestion that the world serves as a sort of external memory (O'Regan, 1992; see also Rensink, 2002), a guidance process tied to preview offset takes on new meaning. Why bother coding into VWM the details of the target when these details are visible and immediately available simply by attending to the target cue? It is only upon the disappearance of this cue that the representation of these details in VWM becomes necessary to guide search. This suggestion is also consistent with “just in time” conceptions of working memory (e.g. Hayhoe et al., 2003). People rarely operate at the limits of their working memory ability when they perform simple routine activities, instead seeking ways to minimize demands on working memory by acquiring visual information just before it is needed by a task (e.g. Ballard et al., 1995; Droll, Hayhoe, Triesch, & Sullivan, 2005; Hayhoe, Bensinger, & Ballard, 1998). With respect to the current task, and assuming that representations start to decay immediately after encoding, waiting until preview offset to begin the feature consolidation process would serve to maximize the amount of time that target information is represented in VWM.

As for how features can be represented in VWM after the target has disappeared from view, one possibility draws on a popular distinction between visual and abstract codes from the iconic memory literature. At least two sub-components of iconic memory have been identified (Irwin & Yeomans, 1986; Irwin & Brown, 1987). One of these is the abstract identity code, believed to be a non-visual coding of stimulus identity and location. Another is the visual analog. This is thought to correspond to the more classic view of iconic memory, in which a high-fidelity representation of the stimulus persists in the form of a visual icon. Importantly, the visual analog was shown to last for ~100–300 ms after stimulus offset regardless of preview duration. If the guiding target representation is constructed from the visual analog at preview offset, this might explain our observation of a boost in guidance only following a brief cue–search delay. This explanation also suggests another answer to the age-old question: What is iconic memory good for (Haber, 1983)? The visual analog component of iconic memory may provide the reservoir of visual detail needed to construct relatively high-fidelity representations of objects in VWM, with the advantage of this being that these representations become time-locked, in an obligatory fashion, to the offset of the stimulus. Of course VWM representations might also be formed at various other times throughout the presentation of a stimulus, with one of these times likely being at stimulus onset. These onset representations would likewise need a few hundred milliseconds of consolidation before they can be used in the rapid detection of objects (e.g., Potter, 1976; Potter & Faulconer, 1975; Vogel et al., 2006; Woodman & Vogel, 2008). However, only a representation formed at object offset maximizes the time that information about that object is available in VWM to mediate a visuo-motor task. As the high-fidelity visual details from the offset representation fade over time, what is left is the more abstract, but enduring, identity code representation, or the related representation posited by object file theory (e.g., Gordon & Irwin, 1996; Gordon, Vollmer, & Frankl, 2008; Hollingworth, 2004; Hollingworth & Henderson, 2002). We speculate that it is this more abstracted representation that mediates the relatively high (and roughly equivalent) levels of search guidance observed before and after the GEE peak.

The short-lived nature of the GEE also provides a clue as to what function this momentary boost in guidance might actually serve. On this matter we can only speculate, but if it is the case that the GEE reflects the operation of an iconic-VWM system designed to code stimulus properties at offset, then this function is likely to extend far beyond the context of visual search. Many everyday tasks require the reacquisition of an object following some brief visual disruption. Tracking is a good example of this; every blink creates an offset relative to the tracked object. The high-fidelity

representation underlying the GEE may be instrumental in our ability to efficiently recover track on these objects following blinks or object occlusions (Scholl & Pylyshyn, 1999). It might also be the case that the GEE is optimized for more naturalistic visuo-motor tasks, ones involving coordinated eye, hand, head, and body movements (Hayhoe & Ballard, 2005). These more complex tasks commonly introduce the need to store visual properties from an object while attention shifts elsewhere to determine how this object is to be used. These sorts of object manipulation tasks seem well suited to benefit from momentary boosts in guidance over the types of delays reported in this study. And of course the GEE is likely to play a role in search, with the importance of this role increasing with the difficulty of the search task. We used a relatively simple search task in this study so as to quantify the GEE unambiguously in terms of search guidance, but our results indicated a guidance boost optimized for delays greater than those typical of initial saccade latencies. This suggests that, under search conditions in which there is no delay between the target cue and search, the GEE may be expressed over the first several eye movements. A logical extension to the present work would be to document the GEE in terms of these non-initial eye movements so as to better understand the scope of its influence on search. More generally, it would be interesting to explore more fully a range of tasks, search and non-search, to learn more about the broader role of guidance in the service of coordinated visuo-motor behavior.

Alternatively, perhaps the GEE did not evolve to fill a specific function, but is rather the useful by-product of the process of creating a perceptual object. This possibility has profound theoretical importance for visual perception and cognition, as it focuses attention on the process of feature consolidation in object creation. The low level coding of real-world objects almost certainly takes place in a very high-dimensional feature space, and it is imperative that we better understand the process used to reduce the dimensionality of this space so as to create more manageable object representations that are optimized to the ongoing task. We believe that the process of feature consolidation in VWM is related to, if not synonymous with, this process of dimensionality reduction. Object file theory has tackled the question of which features or properties survive an early perceptual representation (Gordon & Irwin, 1996, 2000; Gordon et al., 2008; Kahneman & Treisman, 1984; Noles, Scholl, & Mitroff, 2005; Treisman, 1992), but has focused less on the process by which this happens and its time course. It may be that the GEE reflects an intermediate stage in the process of transforming low-level visual information into an abstract object code. In the context of search, a related suggestion is that dimensionality reduction occurs through a process of selectively representing only those features that allow targets to be discriminated from distractors (Zhang, Yang, Samaras, & Zelinsky, 2006). If true, this means that the dimensionality reduction process required by object creation is not universal, but rather is one that can be tailored to how an object is to be used for a specific task. This suggestion is intuitively appealing, and also consistent with our present findings. A consolidation process designed to select discriminative features should selectively remove the noise introduced by a visual mask, as we observed in this study.

An important next step in better understanding the GEE is to more clearly specify the consolidation process underlying its existence. Throughout this paper we referred to consolidation simply as the rapid assembly of features into an efficient guiding representation following the offset of a preview, but this rapid assembly might take either of two forms. One possibility is that the VWM representation itself changes, perhaps from the addition of new visual features allowing for the temporary representation of additional visual details. Another possibility is that the features comprising the target's VWM representation do not change, but rather are weighted in the process of optimizing search guidance.

Assuming that it takes some time to instantiate these weightings, this tuning process might also be considered a form of consolidation, one that is tailored to the specific search task. It may be possible to distinguish between these two possibilities by interleaving a search task with a difficult memory task (e.g., change detection). If the VWM representation changes during consolidation, we should find a delay-dependent boost in memory analogous to the GEE reported here in the context of search. However, if the GEE is a search-specific benefit resulting from rapid feature weighting, we would expect to find it on search trials, but not on memory trials. Further specifying the feature consolidation process underlying the GEE, and determining the task specificity of this process, would seem another important direction for future research.

Acknowledgments

This work was supported by National Institute of Mental Health (NIMH) Grant 2 R01 MH063748-06A1 and National Science Foundation (NSF) Grant IIS-0527585 to G.J.Z. We thank Christian Luhmann, Antonio Freitas, Todd Horowitz, and Geoffrey Woodman, for their thoughtful comments and suggestions in the preparation and review of this paper. We also thank Zainab Karimjee, Rosemary Vespia, Alyssa Fasano, Courtney Feger, Samantha Schmidt, Shreena Bindra, Jon Ryan, Ryan Moore, Arunesh Mittal, Westri Stalder and Todd Dickerson for help with data collection.

References

- Aivar, M. P., Hayhoe, M. M., Chizk, C. L., & Mruczek, R. E. B. (2005). Spatial memory and saccadic targeting in a natural task. *Journal of Vision*, 5(3), 177–193.
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1), 66–80.
- Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, 15(4), 795–801.
- Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, 46(24), 4118–4133.
- Cornelissen, F. W., & Greenlee, M. W. (2000). Visual memory for random block pattern defined by luminance and color contrast. *Vision Research*, 40(3), 287–299.
- Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1416–1438.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433–458.
- Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: Passive, permissive, active or none? *Current Opinion in Neurobiology*, 16(6), 716–722.
- Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, 12(6), 1127–1133.
- Gordon, R. D., & Irwin, D. E. (1996). What's in an object file? Evidence from priming studies. *Perception & Psychophysics*, 58(8), 1260–1277.
- Gordon, R. D., & Irwin, D. E. (2000). The role of physical and conceptual properties in preserving object continuity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26(1), 136–150.
- Gordon, R. D., Vollmer, S. D., & Frankl, M. L. (2008). Object continuity and the transaccadic representation of form. *Perception & Psychophysics*, 70(4), 667–679.
- Haber, R. N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *Behavioral and Brain Sciences*, 6(1), 1–54.
- Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188–194.
- Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, 38(1), 125–137.
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1), 49–63.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 519–537.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113–136.
- Hollingworth, A., & Luck, S. J. (2009). The role of visual working memory (VWM) in the control of gaze during visual search. *Attention, Perception & Psychophysics*, 71(4), 936–949.
- Irwin, D. E., & Brown, J. S. (1987). Tests of a model of informational persistence. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 41(3), 317–338.
- Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance*, 12(3), 343–360.
- Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.
- Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research*, 179, 427–442.
- Meyers, L. S., & Rhoades, R. W. (1978). Visual search of common scenes. *The Quarterly Journal of Experimental Psychology*, 30(2), 297–310.
- Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object file representations. *Perception & Psychophysics*, 67(2), 324–334.
- O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 46(3), 461–488.
- Paivio, A., & Bleasdale, F. (1974). Visual short-term memory: A methodological caveat. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 28(1), 24–31.
- Payne, J. D., Stickgold, R., Swanberg, K., & Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychological Science*, 19(8), 781–788.
- Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, 41(25–26), 3587–3596.
- Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research*, 46(12), 1886–1900.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5), 509–522.
- Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature*, 253(5491), 437–438.
- Rao, R., Zelinsky, G. J., Hayhoe, M., & Ballard, D. (2002). Eye movements in iconic visual search. *Vision Research*, 42(11), 1447–1463.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, 53(1), 245–277.
- Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology*, 62(10), 1904–1914.
- Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, 38(2), 259–290.
- Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, 47(7), 862–875.
- Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), 459–478.
- Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, 5(1), 81–92.
- Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, 9(4), 739–743.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1436–1451.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238.
- Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, 44(12), 1411–1426.
- Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex*, 17(Suppl. 1), 118–124.
- Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review*, 15(1), 223–229.
- Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, 49(16), 2095–2103.
- Zelinsky, G. J. (1999). Precuing target location in a variable set size 'nonsearch' task: Dissociating search-based and interference-based explanations for set size effects. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 875–903.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835.
- Zhang, W., Yang, H., Samaras, D., & Zelinsky, G. J. (2006). A computational model of eye movements during object class detection. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 1609–1616). Cambridge, MA: MIT Press.