Accumulation of visual information across multiple fixations

Yoni Pertzov
The Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel

Galia Avidan
Department of Psychology, Ben Gurion University of the Negev, Beer-Sheva, Israel

Ehud Zohary
The Interdisciplinary Center for Neural Computation and Department of Neurobiology, Hebrew University, Jerusalem, Israel

Humans often redirect their gaze to the same objects within a scene, even without being consciously aware of it. Here, we investigated what type of visual information is accumulated across recurrent fixations on the same object. On each trial, subjects viewed an array comprised of several objects and were subsequently asked to report on various visual aspects of a randomly chosen target object from that array. Memory performance decreased as more fixations were directed to other objects, following the last fixation on the target object (i.e. post-target fixations). In contrast, performance was enhanced with increasing number of fixations on the target object. However, since the number of post-target fixations and the number of target fixations are usually anti-correlated, memory gain may simply reflect fewer post-target fixations, rather than true accumulation of information. To rule this out, we conducted a second experiment, in which the stimulus disappeared immediately after performing a predefined number of target fixations. Additional fixations on the target object resulted in improved memory performance even under these strict conditions. We conclude that, under the present conditions, various aspects of memory monotonically improve with repeated sampling of the same object.

Keywords: eye position, visual memory, fixations, trans-saccadic

Citation: Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision, 9*(10):2, 1–12, http://journalofvision.org/9/10/2/, doi:10.1167/9.10.2.

Introduction

The high resolution fovea is directed to objects of interest using rapid eye movements called saccades. The still periods between saccades are called fixations, in which the retinal image is relatively stable. Although the gist of a scene can be obtained very quickly within a single fixation (Potter, 1976), detailed visual information about a specific object within the scene is acquired by its foveation using saccadic eye movements (Hollingworth & Henderson, 2002; Nelson & Loftus, 1980). As a result, detailed visual processing of scenes is typically a discrete, serial operation in which gaze is sequentially oriented to objects of interest (Zelinsky & Loschky, 2005). Somewhat surprisingly, scene viewing is often characterized by eye movements which are repeatedly directed to the same objects in a seemingly redundant manner (Yarbus, 1967). Refixations occur also during visual search, when the target and distractors maintain a fixed location over time (Gilchrist & Harvey, 2000). Repeated fixations of the same objects were also found in more naturalistic settings. Ballard, Hayhoe, and Pelz (1995) used a computerized task in which they asked participants to copy a model pattern of blocks to a new location on the screen. Interestingly, subjects typically fixated individual blocks in the model pattern more than once, while performing the task. The authors suggested that repeated fixations were utilized to extract different characteristics of the objects (e.g. color information or relative location), at various times, according to the current task requirements. However, we still do not fully understand the benefits of such repeated sampling of an object in the context of scene perception.

One possible utility of repeated fixations may be the accumulation of information about objects and their spatial relationships. The issue of memory across fixations is still a matter of debate. One line of studies suggests that there is no memory for visual information in natural scenes as the world itself acts as an "outside memory" (O'Regan, 1992; O'Regan & Noë, 2001). According to this view, there is no need to store visual information in memory because it can be acquired from the world as needed by shifts of gaze (or covertly by attentional shifts). Along this line, Rensink (2000) proposed that visual memory is limited to the currently attended object in a scene. For an attended object, a coherent visual representation can be maintained across brief disruptions (such as

doi: 10.1167/9.10.2 Received March 8, 2009; published September 4, 2009 ISSN 1534-7362 © ARVO

a saccade, blink, or other brief intervals). However, when attention is withdrawn from an object, the visual object representation disintegrates into its elementary visual features, with no persisting memory (for similar suggestions, see Becker & Pashler, 2002; Wheeler & Treisman, 2002).

This point of view suggests a memory-less behavior even in more naturalistic situations. Accordingly, Ballard et al. (1995) showed that subjects often use a "just-in-time" strategy, deferring acquisition of relevant information until the time the information is needed. However, even in this study, as well as several later ones, it was shown that some memory from previous fixations can be utilized in real world behavior: gaining experience with a specific set of visual stimuli or raising the cost of each gaze shift (by increasing the distance between objects), both lead to enhanced use of the memory trace (Ballard et al., 1995). Similarly, reduction of working memory load lead to greater reliance on previously gathered memory (Droll & Hayhoe, 2007). In fact, many aspects of motor control appear to rely on visual *memory* representations. For example, when leaving a room we easily orient to the door even if it is outside of our field of view, when the movement is initiated. Such planning of movements on the basis of spatial memory information may be more efficient (at least in some cases) than using continuous visual search to locate targets.

These examples demonstrate the importance of information gathered from previous fixations in real world behavior. The present study was designed to investigate whether repetitive fixations on the same object, improve its memory trace, in a cumulative fashion.

Recent studies found evidence for a build-up of visual memory over long presentation times (Melcher, 2001, 2006). In addition, it was demonstrated that the number of direct fixations (or their cumulative time) plays a critical role in this accretion of information. Tatler, Gilchrist, and Land (2005) asked subjects to freely view complex scenes comprised of many objects. The subjects were subsequently asked to report on various visual aspects regarding some of the objects in the scene. They found that position information, but not identity and color, was better remembered with additional target fixations. In another study, Hollingworth and Henderson (2002) showed that change recognition performance improved with the total duration of time spent fixating on a target (across multiple fixations) prior to the occurrence of a change.

Importantly, the latter studies also found evidence for recency effects. Thus, memory of some object properties (Tatler et al., 2005) and change detection performance (Hollingworth & Henderson, 2002) decreased markedly with additional intervening fixations between the last target-object fixation and the end of the trial (termed here: post-target fixations). Critically, however, when subjects fixate the target more often (within a given trial), the time between the last target fixation and the end of trial is likely to be shorter. Therefore, the previously demonstrated memory gain from multiple target fixations may simply reflect fewer post-target fixations, rather than true accumulation of information. We therefore designed Experiment 1 to replicate, in a more controlled environment, the previous findings about accumulation of information. We went a step further by designing a gaze-contingent experiment (#2) which enabled us to isolate the effect of additional target fixations while controlling various confounds.

Methods

Participants

Ten (5 males) and 18 (7 males) naive subjects (ages 19 to 28) took part in one session of Experiments 1 and 2, respectively, in return for course credit. They all gave written informed consent and had normal or corrected-to-normal visual acuity by self-report. Experimental procedures were approved by the ethics committee of the Psychology Department at Ben-Gurion University, Israel.

Experimental design

Experiment 1

Each trial began once the subject pressed a button while fixating a central point. This initiated a drift correction procedure and made the experiment self paced. Next, a white background screen with an array of 8 objects was presented. After 6 seconds of free viewing (see movie of a typical trial) the stimuli disappeared and the subjects were instructed to indicate (using the mouse) which of the 6 object images (two objects: target & foil, each presented in three orientations) was included in the set of stimuli just presented (see Figure 1, multiple choice window), taking into account both object identity and its orientation in 3D. The foil object was randomly picked between all the objects that were not present in the current array. Each object was shown 3 times as the target object (from 3 different view points) and 3 times as a foil. After the subject made his/her choice (correct or incorrect) the correct object in its original orientation appeared and subjects were instructed to actively place it in its original location in the array (using the mouse). The experiment consisted of 150 trials. The order of the trials was randomized for each subject.

Experiment 2

The design was identical to that employed in Experiment 1 except that the presentation time of the stimuli was not fixed. Instead, the stimuli disappeared

Movie 1. A depiction of a typical trial. See Experimental Design for details.

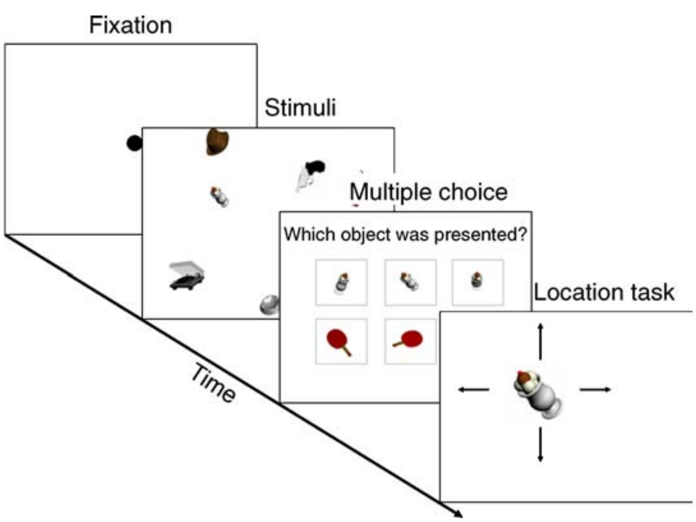

Figure 1. Experimental paradigm. An example of an experimental trial. Following a central fixation point, eight objects were randomly presented on the screen (for 6 seconds in Experiment 1 or a predetermined number of target fixations in Experiment 2) while subjects freely shifted their gaze and inspected the array. Then, a multiple choice screen was presented, in which subjects reported which object was shown earlier and in which orientation. Afterwards, the correct object was presented in its original orientation and the subjects had to indicate its original location within the array by moving the mouse.

immediately after subjects made 1, 2 or 3 fixations on the randomly chosen target object. For each target object, the number of target fixations was counterbalanced across subjects.

Stimuli and experimental settings

The visual stimuli were 150 computer-generated arrays. Each array included 8 different objects located in random positions on the screen (excluding a 4 degree perimeter around the center of screen, see Figure 2). The objects were selected from a pool of 50 distinctive objects rendered in 3 different orientations (3D studio MAX; Autodesk, Inc, Montreal). The objects' identities as well as their specific orientation were randomly assigned in each trial. Each combination of object and orientation (150) appeared in 8 different arrays, once as the target object and seven times as a "distractor." The images were displayed on a 19-inch CRT monitor (Graphics Series G90fB, View Sonic, Los Angeles, USA), at a resolution of 1024 × 768 pixels with a refresh rate of 100 Hz. The stimuli were within 34.3 × 25.8 degrees on the horizontal and vertical axes, respectively. Further descriptions of the stimuli are provided above in the "Experimental design" section.

Eye tracking

The experiments were conducted in a dimly lit room, subjects sat in front of a computer screen while their head was positioned in a chinrest. Subjects' eyes were located 60 cm from the computer screen. A video-based desk-mounted eye tracker (Eye Link1000, SR Research, Ontario, Canada) with a sampling rate of 1000 Hz was used for recording eye movements. We used built-in programs provided with the eye tracker for calibration and validation purposes (9 points in a random sequence). All the data analyzed in the present paper were obtained from recordings with an average absolute global validation error of less than 1 degree.

Each trial began with the presentation of a fixation point at the center of the screen. The subjects triggered (by a key press) the stimulus display when they were ready while fixating this point. The data obtained during this control fixation period were used to correct for slow drifts of the eye tracker. If drift errors were high (more than 2 degrees) a new calibration protocol was initiated. After every 50 trials subjects had a break which was followed by an additional calibration procedure.

Fixation points and saccades were defined using the following procedure: For each data sample, the built-in event parser of the eye tracker computed instantaneous velocity and acceleration. The saccade detector became "active" if these two values crossed predefined thresholds (30 deg/s and 8000 deg/s^2 respectively). A saccade was defined as the period of time in which the saccade detector remained active (for at least 2 samples in a sequence) until one of the two parameters was below threshold (signaling the saccade detector inactivity) and remained so for at least 20 msec. The episodes between saccades were defined as fixation events. The resulting scanning patterns were plotted and visually inspected to check that they produced adequate parsing of the eye-position samples to saccades and fixations.

In Experiment 1, the parsing of the gaze samples to fixations and saccades was calculated offline. Experiment 2, however, required online parsing to assure that the display ends immediately after a predefined number of fixations on the target. In this case, fixation was considered as a contiguous period (more than 50 msec) in which the saccade detector threshold was not crossed. The display stimuli disappeared immediately (~10 msec delay) after the last predetermined fixation on target (1st, 2nd or 3rd fixation) ended (by crossing of the saccade detector threshold). The two procedures were shown to generate similar results when applied on the same data set (see supplementary Figure S4).

Analysis of fixations

A target perimeter (TP) was defined around each target object. This was defined as a circle with radius of ~3.5 degrees surrounding the object. Each TP exclusively included a single target object. We counted the number of fixations within the TP per each trial (target fixations), and the number of post-target fixations (i.e. the number of fixations after the last fixation on the target, until the end of the trial, see Figure 2). We then analyzed the subjects' memory performance with respect to these factors.

Measures of memory performance

On each trial, one object was randomly chosen to be the target object. Two tasks were used to measure acquired knowledge about the target object. First, the subjects chose which of the 6 options (two objects, each presented in three orientations) was included in the set of stimuli just presented. This allowed us to assess memory of the object's identity as well as its orientation (as each object was presented from three different viewpoints). Note, however that choosing the right object might be mediated by partial information (e.g. color, shape, etc), rather than a full representation of the object identity. Next, subjects marked the position of the target object. The Cartesian distance from the object's true location was registered as the location error. We calculated the "chance" location error separately for each subject. This was the averaged Cartesian distance between the location of the target objects and the shuffled reported target locations. Because this procedure uses the true location reports of the subjects

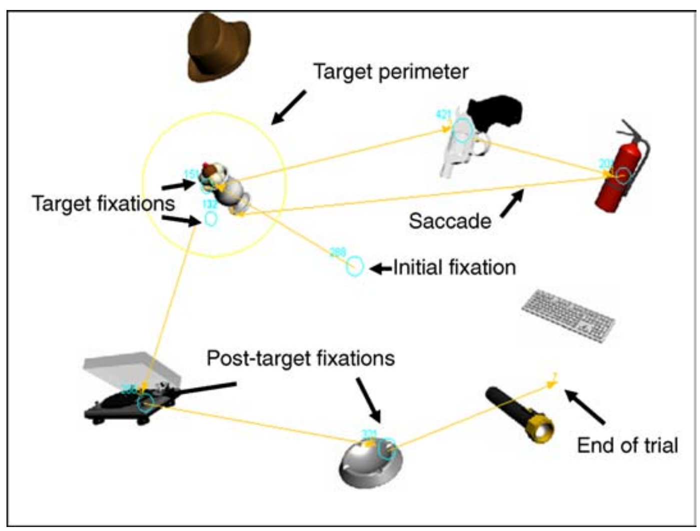

Figure 2. Procedure for parsing experimental data. An example trial (with fewer number of fixations and saccades than in reality, for the sake of clarity). First, gaze position data is parsed to fixations (circles in cyan, adjacent number indicates duration of fixation) and saccades (orange lines). Then, the numbers of target fixations (two in this case) and post-target fixations (i.e. the number of fixations after the last fixation on the target, until the end of the trial) are counted (two in this example). Reproduced with permission. Copyright The Hebrew University of Jerusalem.

(in a shuffled temporal trials' order), it controls for reporting biases, such as the subjects' tendency to report the locations near the center of the screen. Finally, we calculated the percent of improvement in locating performance, relative to chance. In that way, similarly to identification and orientation scores, better performance is represented as higher values.

Results

The goal of our experimental design was to investigate the influence of multiple direct fixations on the memory trace of natural objects; we therefore presented subjects with 150 arrays, each including 8 different objects, "floating" in random positions. In each trial, one object was preselected as the target (but this was obviously unknown to the subject). The subjects were required to identify the target object, judge its orientation and report its original location in the array (see movie of a typical trial). Performance in these memory tests was analyzed in relation to the subjects' eye-scanning patterns.

Experiment 1

First, we sorted out trials according to the number of fixations on the target object (see Figure 2). Note that fixations on target objects could either be consecutive or separated in time. Both cases were included in the analysis below. Next, we computed the average task performance with respect to the number of fixations on the target object (Figure 3).

As expected, our results show a clear improvement of memory performance with greater number of target fixations. One way repeated measures ANOVA on memory performance with linear trend analysis for target fixations (0, 1, 2, 3, and 4 or more) was conducted to quantify the effects. ANOVA of the target identity performance across the four possible levels of fixations revealed a significant effect [$F(4,36) = 12.4$, $p < 0.0001$] and a linear trend [$\eta_p^2 = 0.74$, $p < 0.0001$] such that additional fixations yielded better target identification. Importantly, performance was significantly better than chance [two tailed t-test, $p < 0.02$] even during trials with zero fixations on the target; this implies that some information on the identity of an object can be gathered from extra-foveal vision. Note that this information could be partial (e.g. specific features of the object such as color or general shape) rather than a full representation of the object identity. ANOVA of the orientation performance revealed a marginally significant effect of the number of target fixations [$F(4,36) = 2.63$, $p = 0.05$] and a linear trend [$\eta_p^2 = 0.44$, $p < 0.03$] such that additional fixations yielded better viewpoint reports. The performance in trials

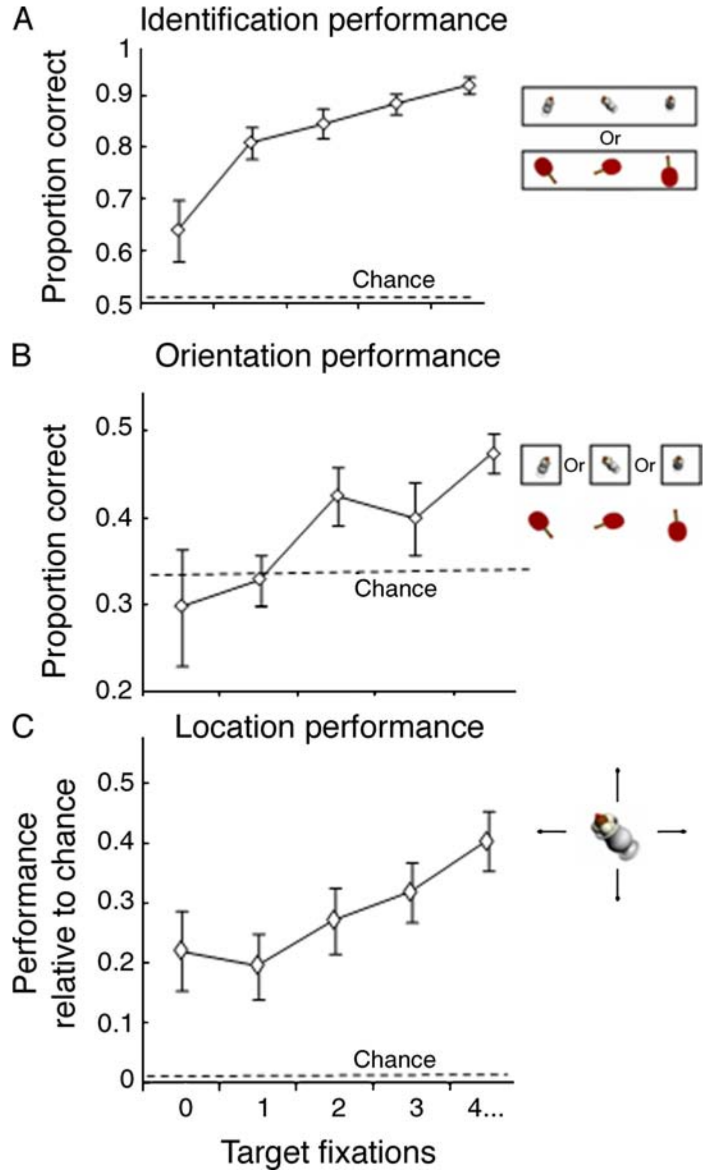

Figure 3. Memory performances with respect to the number of fixations on target. Performance scores (proportion correct) for the three parameters tested, as a function of the number of fixations on the target object. (A) Target object identification (chance = 0.5). (B) Target object viewpoint judgment (chance = 0.33). (C) Target object localization (chance = 0). Error bars denote standard error of the mean across subjects ($N = 10$).

with zero fixations on the target is not significantly [two tailed t-test, $p > 0.63$] better than chance (0.33) suggesting that information on the exact viewpoint of an object is not available from extra-foveal vision. Finally, ANOVA of location performance revealed a significant effect of the number of target fixations [$F(4,36) = 6.3$, $p < 0.0006$] and a linear trend [$\eta_p^2 = 0.7$, $p < 0.002$] such that additional fixations yielded smaller location error. The performance in trials with zero fixations on target was significantly better than chance [two tailed t-test, $p < 0.005$].

We conclude that performance in all three memory tasks is enhanced in trials with additional fixations on the target object. Furthermore, some information about the target identity and its position in space can be obtained from peripheral vision (though correct judgment of its orientation requires direct fixation).

Next, we analyzed memory performance with respect to the number of post-target fixations (fixations between the last fixation on the target and the end of the presentation of the stimulus array).

The results of target identification and location (but not target orientation) indicate a clear decrease in memory performance as the number of post-target fixations increases (see Figure 4). One way repeated measures ANOVAs with linear trend analysis were conducted to quantify the effects. ANOVA of target identification across the five categories of post-target fixations (0, 1, 2, 3 and 4 or more) revealed a significant effect [$F(4,36) = 3.2$, $p < 0.024$] and a linear trend [$\eta_p^2 = 0.74$, $p < 0.001$] such that additional post-target fixations resulted in higher location errors. Interestingly, ANOVA of orientation performance did not reveal a significant effect of post-target fixations [$F(4,36) = 1$, $p > 0.39$]. Thus, it appears that the knowledge retained on the orientation of the object is not affected by the number of post-target fixations. ANOVA of location performance revealed a significant effect of the number of post-target fixations [$F(4,36) = 10.2$, $p < 0.00001$] and a linear trend [$\eta_p^2 = 0.8$, $p < 0.0003$] such that additional post-target fixations led to higher location errors.

In the results described above, we grouped the trials according to the number of target and post-target fixations. One could also use an alternative method, binning the trials according to the total time that gaze was directed to the target object. The results obtained using this second analysis were generally the same as those obtained when the discrete number of fixations were the dependent measure (see supplementary Figure S2), generally replicating previous results (Tatler et al., 2005). The similarity between the results obtained using these two approaches is not surprising given that the number of fixations and total fixation time are likely to be highly correlated. Using the number of target fixations as our dependent measure has the advantage that it eliminates the need to determine the appropriate bin size in the time domain. It is also better suited for the gaze contingent method used in Experiment 2 (see below).

To summarize the results presented so far, direct fixations are clearly important for memory performance. Additional fixations on the target object lead to better performance in all three memory parameters tested here. In addition, fixations that occur after the last fixation on the target (post-target fixations) seem to have a differential effect on the persistence of each of the memory parameters. Specifically, memory for target identity and location (but not target orientation) decreased monotonically as more post-target fixations occurred. This suggests a

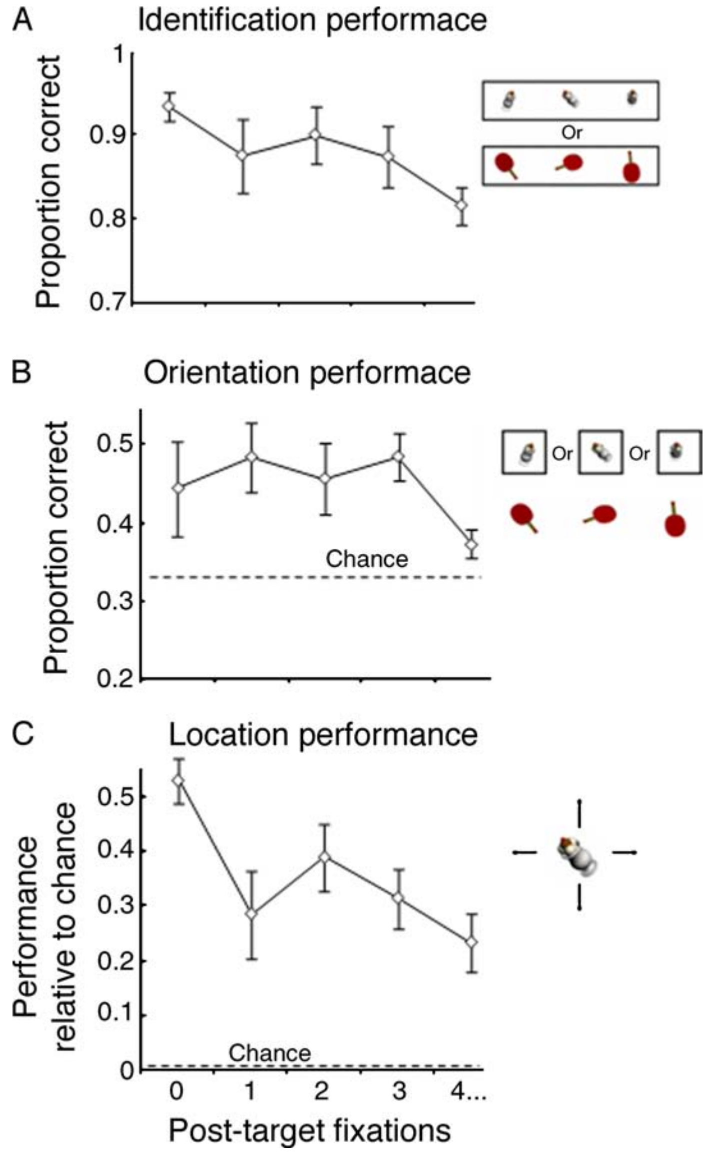

Figure 4. Memory performance with respect to the number of post-target fixations. Average performance scores for each of the three types of information tested in Experiment 1 as a function of the number of post-target fixations. (A) Object identification performance (chance = 0.5). (B) Object viewpoint performance (chance = 0.33). (C) Object localization (chance = 0). Error bars denote standard error across subjects (*N* = 18). Note that chance level is not shown on panel A as it is below the range of the scale used in this panel.

selective decay process of the obtained memory traces with post-target fixations (or time). One major caveat in interpreting our results is that on a trial-by-trial basis the number of fixations on the target is anti-correlated with the number of post-target fixations [$r = -0.31$; averaged across subjects]. This can be intuitively understood: when one is making more fixations on the target, the last saccade on the target is likely to occur closer to the end of the trial thereby reducing the number of post-target fixations. The correlation (between the number of target fixations and the number of post-target fixations) therefore precludes any conclusion regarding the effect of only one of these factors. For example, memory gain from multiple target fixations may simply reflect fewer post-target fixations, rather than true accumulation of information. We used two approaches to deal with this possible confound: The first was to carry out an additional, more sophisticated statistical analysis on the current results based on part correlation. The second approach was to conduct another experiment, in which the number of target fixations was predefined. Both approaches are described in detail below.

Part correlation analysis

We used part correlation analysis which allows us to examine the relationship between two variables after the effect of the third variable (on both) has been factored out. Specifically, we examined whether memory performance would still be correlated to the number of target fixations even when post-target fixations are controlled (i.e. factored out). This analysis requires a within-subject analysis as the correlation between post-target and target fixations is seen only on a trial-by-trial basis.

A trial-by-trial analysis, at the individual subject level, also revealed similar effects to those described above (albeit weaker compared to the group analysis). On a trial-by-trial basis, the number of fixations on target is correlated to memory performance [r = 0.19, 0.14, 0.16 average across subjects for identification, orientation and location performance, respectively] and significantly different from zero [t-test after fisher transformation; $p < 0.0001$ for all memory tests]. The number of post-target fixations is also correlated to memory performance [r = −0.12, −0.1, −0.17 average across subjects for identification, orientation and location performance, respectively] and significantly different than zero [t-test after fisher transformation; $p < 0.0001$ $p < 0.03$ $p < 0.0002$ for identification, orientation and location, respectively].

Importantly, when performing *part* correlation, between the number of fixations on the target and memory performance, in which the effect of post-target fixations is removed, correlation prevails [r = 0.16, 0.13, 0.12; average across subjects for identification, orientation and location performance, respectively] and is significantly different from zero [t-test after fisher transformation $p < 0.001$ for all memory tests].

Experiment 2

The second approach to tackle the issue of anti-correlation between post-target and target fixations was by designing a new experiment (#2) in which the number of post-target fixations was experimentally controlled. Thus the number of target fixations was predefined using a gaze contingent presentation method. The design was similar to Experiment 1, however, in this second experiment, memory tests were conducted after subjects performed a predefined number of target fixations. An online process checked for target fixations and eliminated the object array immediately after 1, 2 or 3 fixations on the target object.

We sorted the trials of each subject according to the number of fixations on target. Then, we computed the average performance (across subjects) for each group of trials (Figure 5).

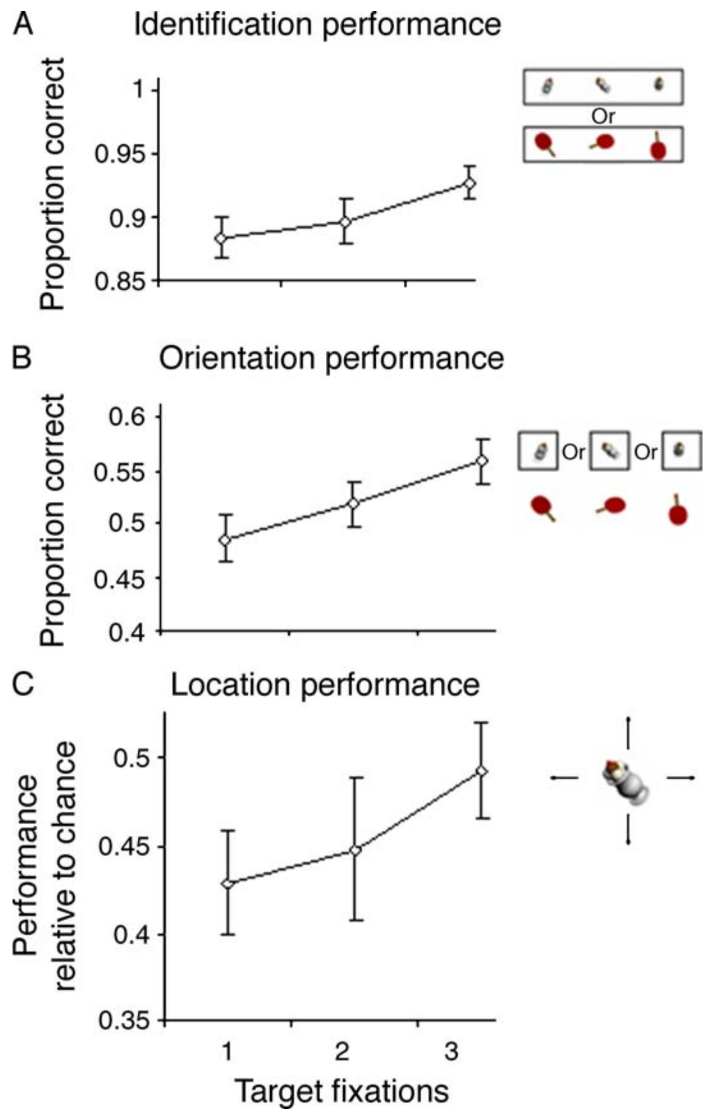

Figure 5. Memory performance with respect to the number of fixations on target, controlled for the number of post-target fixations. Performance scores (N = 18) for each of the three types of information tested in Experiment 2 as a function of the number of fixations on the target object. (A) Object identification performance (chance = 0.5). (B) Object viewpoint performance (chance = 0.33). (C) Object localization (chance = 0). Note that chance levels are not shown on the graphs as they are below the range of the scale used here.

In all three memory tasks, performance was enhanced when subjects made additional fixations on the target object. A one way repeated-measures ANOVA with linear trend analysis was conducted to quantify the effects. ANOVA of target identification performance revealed a significant effect of the number of fixations on the target [$F(2,34) = 5.5$, $p < 0.01$] and a linear trend [$\eta_p^2 = 0.32$, $p < 0.02$] such that additional fixations yielded better target identification. ANOVA of target orientation performance also revealed a significant effect of target fixations [$F(2,34) = 4.24$, $p < 0.03$] and a linear trend [$\eta_p^2 = 0.29$, $p < 0.02$] such that additional fixations yielded better viewpoint reports. ANOVA of location performance revealed a marginally significant effect of target fixations [$F(2,34) = 2.9$, $p = 0.07$] and a linear trend [$\eta_p^2 = 0.33$, $p < 0.01$] such that additional fixations resulted in better target localization. These results further strengthen the notion that information, as measured in all three types of memory tests, is accumulated as a function of the number of target fixations.

Discussion

Summary

Subjects freely viewed arrays of several objects, before answering questions about one selected target object. In the first experiment, in which the scanning time of the array was predetermined, performance was found to improve with the number of fixations on the target item. The number of post-target fixations, following the last fixation on the target item, was also found to be correlated with performance: more post-target fixations led to decreased performance in the target identification and localization measures. A second experiment was designed to overcome the inherent correlation between the two above factors, by extinguishing the visual array immediately after a predefined number of fixations on the target object. This experiment provided further evidence that performance in all memory tasks improves with increasing number of target fixations.

In the real world, objects never appear in isolation; they co-vary with other objects and particular environments, providing a rich source of contextual associations to be exploited by the visual system. This contextual information clearly influences scanning patterns and visual memory (Oliva & Torralba, 2007; Tatler & Melcher, 2007). In our experiment, we made an effort to abolish any influence of prior or contextual knowledge from affecting task performance (e.g. a vase is likely to be placed on the table). We therefore used an array of isolated objects which is clearly less naturalistic but avoids the major "contaminating" effects of prior expectations on the objects' location and orientation.

In this discussion, we first rule out the possibility that our results may have been due to some other confounding effects, and then discuss our results regarding the specific characteristics of memory gathered from free viewing.

Ruling out possible confounding effects

Improved performance with more target fixations may be due to fewer post-target fixations

One major goal of the present study was to investigate the functional benefits of repeatedly refixating the same object during free viewing. Two previous studies (Hollingworth & Henderson, 2002; Tatler et al., 2005) suggested that under free viewing conditions, refixation on the same object leads to accumulation of information about some of its characteristics. Both studies also found that some aspects of performance deteriorated with the number of fixations (or time) following the last fixation on the target (post-target fixations). In essence, we replicate and extend these results using a different methodology in Experiment 1. However, their conclusion about the accumulation of information across repeated fixations was somewhat premature as an important confound was neglected. Specifically, when more fixations are directed to the target, the last fixation on the target is likely to occur later during the trial. This, in turn, leads to fewer post-target fixations. Indeed, our results reveal significant negative correlation between the number of "target fixations" and "post-target fixations" across trials. This confounding effect (Figure 6A), may well explain an apparent direct causal relationship between the number of target re-fixations and memory performance (see Figure 6B), as the effect of additional target fixations on performance could be attributed to fewer post-target fixations.

We used two methods to deal with this caveat. The first was to employ part correlation statistical analysis on the data obtained from Experiment 1. Specifically, we found that memory performance was still correlated to the number of target fixations even after the effect due to the number of post-target fixations was factored out (see Results section: Part correlation analysis). The second, perhaps more compelling piece of evidence for true accumulation of information across target re-fixations comes from Experiment 2: Here, in spite of the fact that no post-target fixations were present in any of the trials, performance was enhanced in all tasks when additional fixations were directed to the target. The two experiments together, provide substantial evidence for accumulation of information across fixations.

Objects differ in the degree to which they attract attention

For various reasons, some objects attract more attention than others. For example, a red object embedded in an array of gray objects, will automatically attract attention

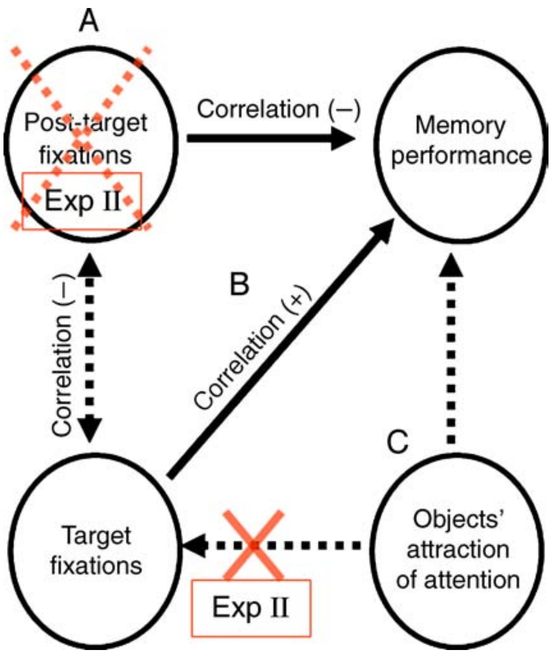

Figure 6. Direct and indirect linkages between the number of target fixations and memory performance. The correlation between the number of fixations on the target and memory performance could be a result of a common factor (C) rather than a direct causal influence (B). In Experiment 2, by predefining the number of target fixations on each trial, target fixation was made an independent variable. Experiment 2 also eliminated any post-target fixations—therefore disabling the dependency between post-target and target fixations (A).

and would therefore be a more likely target for fixation. Accordingly, objects with distinct image properties such as greater luminance, chromatic contrast, or motion are more likely to be fixated (Itti, 2005; Itti & Koch, 2000; Tatler, Baddeley, & Vincent, 2006; but see Foulsham & Underwood, 2008; Tatler & Melcher, 2007). The attraction of gaze to specific objects could also be mediated by various high-level factors, such as object recognition (Einhauser, Spain, & Perona, 2008) and other cognitive processes (Henderson, Brockmole, Castelhano, & Mack, 2007).

Salient objects are also more likely to be remembered and selected in free recall (Einhauser et al., 2008) and categorization tasks (Elazary & Itti, 2008). This raises a potential problem: If salient objects are more memorable, *even without* attracting gaze towards them, the correlation between the number of fixations on the target and memory performance could be a result of a common factor i.e. object saliency (see Figure 6C) rather than reflecting a direct causal influence between the two. We addressed this confound in two ways. First, by predetermining the number of target fixations on each trial in Experiment 2, we assured that these were no longer influenced by the object's saliency. In other words, object saliency is still expected to influence memory performance, but it no longer affects the number of target fixations, as this number is experimentally controlled. This manipulation ruled out the possibility that object saliency (or any differences among objects in terms of "attention capture") is independently affecting both target fixations and memory performance, and therefore contributing to the relation between the number of fixations and memory performance in Experiment 2.

We also used our gaze-contingent method (Experiment 2) to directly explore whether more salient objects would be better remembered, irrespective of the number of fixations directed on them. To that end, we divided the objects into two groups of saliency (high and low) according to the number of fixations they attracted in Experiment 1 (using a fixed duration). The logic behind this grouping was the notion that the more salient objects would be fixated more often. We then compared memory performance for the two object groups in Experiment 2 (when the number of target fixations was controlled; see supplementary Figure S3). Our results indicate that more "attractive" objects do not seem to be more memorable than less "attractive" objects, when the number of target object fixations is equated. Therefore, better memory of salient objects seems to be mediated by more frequent fixations, rather than by object saliency per se.

Primacy effects

The fact that objects differ, in the degree to which they attract attention, could possibly confound our results in yet another way. More salient objects would probably be fixated at the beginning of the trial. In addition, objects which were fixated at the beginning of the trial are expected to be remembered more accurately—an effect called "Primacy effect" in serial memory studies (Wright, Santiago, Sands, Kendrick, & Cook, 1985). This effect could possibly influence the results in the same fashion as the number of post-target fixations affects performance. In order to investigate this possibility, we analyzed memory performance in Experiment 1 as a function of the number of fixations *before* the first time the object was directly fixated. Whether the target object was fixated for the first time after 1, 2, 3 and more-than-4 fixations revealed no effect on any of the memory scores ($F(3,27) < 0.7$, $p > 0.6$ for identification, orientation and location scores, see supplementary Figure S1). Primacy effects, therefore, do not seem to influence memory performance under our experimental settings. Note that previous studies of object-memory under free viewing also did not report any "primacy effects" (Tatler et al., 2005; Zelinsky & Loschky, 2005).

The possible effect of additional fixations on non-target objects

Experiment 2 handles the two first confounds mentioned above. However, it fails to address another one. In

Experiment 2, trials with three target fixations are longer than trials with only one target fixation. Therefore, they are also likely to include more fixations on *non-target* objects. The additional fixations on non-target objects might improve the memory performance regarding the target object, especially in the localization task. We must take this into account before proposing a true accumulation of information across repeated saccades. Note, however, that in Experiment 1 the trial duration was fixed and therefore the overall number of fixations was approximately equal in the different trials. Our additional analysis of Experiment 1 using a trial-based part correlation between target fixation and memory performance (when factoring out the effect of post-target fixations) showed that this correlation is still significantly positive (though smaller). Therefore, from the two experiments together, we can conclude that multiple fixations on the same object lead to true accumulation of information about its features (e.g. identity, location and orientation).

Extrafoveal information

Trials with zero fixations on the target (Experiment 1) deserve special attention, as in such trials the fovea was never aimed directly at the object of interest. In such cases, any information regarding the target object comes only from extrafoveal vision. While visual acuity clearly deteriorates from foveal to extrafoveal vision, it was found to be sufficient to acquire some information about the object identity (Thorpe, Gegenfurtner, Fabre-Thorpe, & Bulthoff, 2001), mainly through low spatial frequency channels (Pointer & Hess, 1989). Indeed, several studies found evidence for extrafoveal preview benefits for object recognition (Henderson & Anes, 1994; Henderson, Pollatsek, & Rayner, 1987; Pollatsek, Rayner, & Collins, 1984; Pollatsek, Rayner, & Henderson, 1990). Specifically, when subjects were required to name a centrally presented object, naming latencies were shorter when an extrafoveal preview of the same object was presented prior to the central presentation. Note, however, that subjects were not required to process any foveal information during the extrafoveal preview. It is reasonable that extrafoveal information extraction during free-viewing is much more limited, since attentional resources are captured by various other tasks such as foveal vision and target selection. Only a handful of studies were published on memory performance gained through extrafoveal vision during free-viewing. Tatler et al. (2005) found that object identity (but not its shape, color or position) can be remembered even without direct fixation. Our study similarly showed that some basic aspects of the object's identity (possibly basic shape or color) can be gathered through extrafoveal vision. However, more detailed information, such as the object orientation, requires foveal vision. The only discrepancy between our study and Tatler et al. (2005) is that they found chance performance for reporting the position of non-fixated objects, whereas the present study showed above-chance performance. This could possibly be explained by the different reporting measures used. While Tatler et al. (2005) used a paper questionnaire with only four possible answers; we used a more flexible approach in which subjects actively registered the object position by moving the mouse, allowing for better positional resolution. Another possible cause for this apparent discrepancy is the different method used for calculating chance level. To estimate this measure, Tatler et al. (2005) used another group of subjects that filled the questionnaire without ever seeing the scenes. In contrast, we used the actual object positions indicated by the same subject (after shuffling the order of the trials) to quantify chance level. Our approach may provide a more sensitive measure of performance with respect to chance, as the baseline is not calculated on a different group of subjects.

Our findings, together with those of Tatler et al. (2005), nicely fit with the general account of active vision (Findlay & Gilchrist, 2003). Accordingly, information about the gist of the scene (such as some vague sense of the objects' identity and their general location) is extracted from extrafoveal vision. In natural viewing, this information is used to guide our eyes to selected objects in order to further gather fine-detailed information, such as its exact orientation, which requires direct fixation.

Orientation specificity

Importantly, unlike object location and identification the object orientation memory trace seems to be less influenced by post-target fixations. Furthermore, orientation information (again, unlike identification and location information) seems to be extracted only from foveal vision. These differences could possibly be explained by the distinctive type of visual information needed for the different memory measures: When gaze (and attention) is oriented to a local object in a scene, low-level visual processing is highly biased towards the object falling on the fovea, and fades away with the next saccade (see Henderson & Hollingworth, 2003; Hollingworth, Richard, & Luck, 2008; Irwin, Yantis, & Jonides, 1983; O'Regan & Levy-Schoen, 1983). However, visual processing also leads to the construction of more abstract representations at higher levels of analysis. These may include a visual description of the attended object (or a semantic one), obtained from its low-level properties. Information on the identity and general location of the object might rely on this higher level analysis. These higher level visual representations, by virtue of being invariant to natural changes in the low-level features of objects (for instance, due to lighting conditions), are inherently insensitive to the specific visual details of an object (such as the exact viewpoint at which the object was observed). Therefore, encoding of such properties might require more processing resources. This extra processing might, in turn, lead to better stability of the viewpoint memory trace.

Concluding remarks

We conclude that some information about objects, such as their fine details and exact location, fades away when fixations are directed elsewhere. When gaze (and attention) is redirected to an object, the task relevant information is further accumulated. Some information about an object's location and its general identity can be extracted from extrafoveal vision; This information, in turn, could be used to guide our eyes to gather more detailed information on the objects' specific characteristics (such as the exact viewpoint in which the object is presented) which requires foveal vision, and seems to be more stable across time.

Acknowledgments

We would like to thank Ido Peleg for his help with data acquisition. We thank Michal Jacob for her help in the experimental settings.

This work was supported by the National Institute for Psychobiology in Israel (NIPI) grant 2-2008-09 to GA and an Israel Science Foundation (ISF) grant to EZ.

Commercial relationships: none.
Corresponding author: Yoni Pertzov.
Email: pertzov@gmail.com.
Address: Interdisciplinary Center for Neural Computation and the Department of Neurobiology, Hebrew University, Jerusalem, Israel 91904.

References

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7,* 66–80.

Becker, M. W., & Pashler, H. (2002). Volatile visual representations: Failing to detect changes in recently processed information. *Psychonomic Bulletin & Review, 9,* 744–750. [PubMed] [Article]

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1352–1365. [PubMed] [Article]

Einhauser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1–26, http://journalofvision.org/8/14/18/, doi:10.1167/8.14.18. [PubMed] [Article]

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision, 8*(3):3, 1–15, http://journalofvision.org/8/3/3/, doi:10.1167/8.3.3. [Article]

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*(2):6, 1–17, http://journalofvision.org/8/2/6/, doi:10.1167/8.2.6. [PubMed] [Article]

Gilchrist, I. D., & Harvey, M. (2000). Refixation frequency and memory mechanisms in visual search. *Current Biology, 10,* 1209–1212. [PubMed]

Henderson, J. M., & Anes, M. D. (1994). Roles of object-file review and type priming in visual identification within and across eye fixations. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 826–839. [PubMed]

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel MHF, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford, England: Elsevier.

Henderson, J. M., & Hollingworth, A. (2003). Global transsaccadic change blindness during scene perception. *Psychology Science, 14,* 493–497. [PubMed]

Henderson, J. M., Pollatsek, A., & Rayner, K. (1987). Effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 449–463. [PubMed]

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,* 113–136.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General, 137,* 163–181. [PubMed]

Irwin, D. E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics, 34,* 49–57. [PubMed]

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12,* 1093–1123.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506. [PubMed]

Melcher, D. (2001). Persistence of visual memory for scenes. *Nature, 412,* 401. [PubMed]

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision, 6*(1):2, 8–17, http://journalofvision.org/6/1/2/, doi:10.1167/6.1.2. [PubMed] [Article]

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology, 6,* 391–399. [PubMed]

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends Cognitive Science, 11,* 520–527. [PubMed]

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46,* 461–488. [PubMed]

O'Regan, J. K., & Levy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research, 23,* 765–768. [PubMed]

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral Brain Science, 24,* 939–973; discussion 973–1031. [PubMed]

Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research, 29,* 1133–1151. [PubMed]

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113,* 426–442. [PubMed]

Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 199–210. [PubMed]

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology, 2,* 509–522. [PubMed]

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research, 40,* 1469–1487. [PubMed]

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46,* 1857–1862. [PubMed]

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology: A, Human Experimental Psychology, 58,* 931–960. [PubMed]

Tatler, B. W., & Melcher, D. (2007). Pictures in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception, 36,* 1715–1729. [PubMed]

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bulthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *Europe Journal of Neuroscience, 14,* 869–876. [PubMed]

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131,* 48–64. [PubMed]

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys, and people. *Science, 229,* 287–289. [PubMed]

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Zelinsky, G. J., & Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. *Perception & Psychophysics, 67,* 676–690. [PubMed] [Article]