Visual scene memory and the guidance of saccadic eye movements

David Melcher *, Eileen Kowler

Department of Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA

Received 3 January 2001; received in revised form 12 June 2001

Abstract

An unresolved question is how much information can be remembered from visual scenes when they are inspected by saccadic eye movements. Subjects used saccadic eye movements to scan a computer-generated scene, and afterwards, recalled as many objects as they could. Scene memory was quite good: it improved with display duration, it persisted over time long after the display was removed, and it continued to accumulate with additional viewings of the same display (see also Melcher, Nature (2001) in press; Doctoral dissertation (2001)). The occurrence of saccadic eye movements was important to ensure good recall performance, even though subjects often recalled non-fixated objects. Inter-saccadic intervals increased with display duration, showing an influence of duration on global scanning strategy. The choice of saccadic target was predicted by a Random Selection with Distance Weighting (RSDW) model, in which the target for each saccade is selected at random from all available objects, weighted according to distance from fixation, regardless of which objects had previously been fixated. The results show that the visual memory that was reflected in the recall reports was not utilized for the immediate decision about where to look in the scene. Visual memory can be excellent, but it is not always reflected in oculomotor measures, perhaps because the cost of rapid on-line memory retrieval is too great. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Visual memory; Saccadic eye movements; Memory retrieval

1. Introduction

A fundamental problem in the study of human vision is how information accumulates across separate glances and over time into a lasting representation of the whole visual scene. Accumulation is important because only a handful of items can be remembered from any single fixation (Sperling, 1960). The information available for cognition and behavior must build up over time, or else humans would be trapped in the eternal present. When looking around a scene, information accumulated over fixations must serve both the immediate needs of action, such as selecting the target for saccadic eye movements, and long-range cognitive goals, such as learning about the environment. The purpose of this study is to examine the way that information acquired from individual fixations is used both for selecting saccadic targets, and also for building a more enduring representation of the entire scene.

Whether it has been studied in the context of immediate action, or in terms of the learning about the environment, there is a surprising lack of consensus about the basic properties of scene memory. Many early studies of scene recognition reported excellent memory performance. The overall ‘gist’ of a scene, for example, can be recognized after a single glance (Potter, 1976; Biederman, 1972; Kundel & Nodine, 1975; Loftus, 1972) and can even lead subjects to incorrectly ‘recognize’ semantically consistent objects from scenes that were not actually there (Biederman, 1972; Intraub & Richardson, 1989; Miller & Gazzaniga, 1998). Thus, people appear to gather semantic information (Gordon & Irwin, 2000) about selected items, and semantic information remains in memory across separate fixations (Hollingsworth, Schrock, & Henderson, in press).

On the other hand, studies using different approaches have led to very different conclusions about visual memory. For example, recent studies of eye movements during natural tasks have concluded that memory for
scene information can be quite limited. Analyses of eye movement patterns in tasks such as solving geometry problems (Epelboim & Suppes, 2001), copying a block pattern (Ballard, Hayhoe, & Pelz, 1995), and recalling object location (Zelinsky & Loschky, 1998), have shown that people use saccadic eye movements to repeatedly refer back to the physical scene. The frequent occurrence of these revisiting eye movements was attributed to poor memory for the contents of the scene.

Analysis of eye movement patterns in other tasks, however, suggests that maybe memory limits are not so severe, at least as far as object location is concerned. In well-practiced activities, such as making tea or sandwiches (Land & Furneaux, 1997; Land, 2000; Land & Hayhoe, in press), or repeatedly tapping the same set of objects (Epelboim et al., 1995), subjects look directly at task-relevant locations without needing extra eye movements to search for the relevant objects in the scene. Visual memory for at least part of the scene must have been good enough to allow the important objects to be located quickly.

Perhaps the discrepancies among the eye movement studies are to be expected because it is not always easy to infer cognitive organization based solely on eye movement patterns (Viviani, 1990). Eye movement patterns can reflect several concurrent task goals and may even be programmed ahead of time as an entire sequence (Zingale & Kowler, 1987). During visual search, for example, subjects may re-fixate previously viewed distractors, and often continue looking around the display even after they have found the target — results that illustrate discrepancies between eye movements and cognition (Engel, 1977; Gould, 1973; Hooge & Erkelens, 1998; Zelinsky, 1996). In the block-copying (Ballard et al., 1995) and geometry (Epelboim & Suppes, 2001) experiments, where frequent re-visits to important objects were reported, it is not necessarily the case that the revisiting implies poor memory capacity. Memory may have been underutilized in these tasks, perhaps because attention to the task itself drew cognitive resources away from memory formation, or because subjects knew they could easily look at the display to confirm the contents whenever they felt uncertain (Oliva, in press). It would seem that the capacity to remember visual objects might be better reflected in the prior eye movement studies using familiar or well-practiced tasks (Land & Furneaux, 1997; Land, 2000; Epelboim et al., 1995), but caution is needed here as well, because extensive practice could represent a special case, rather than visual memory as it normally operates in everyday situations.

Analogous complexities in interpretation apply to psychophysical studies of the phenomenon known as ‘change blindness’. These studies have focused on the limitations of visual memory as inferred by failure to notice changes to items in a complex scene that occur during actual or simulated saccades (Rensink, O'Regan, & Clar, 1997). Changes to items of ‘central interest’ are usually detected, which has been interpreted as evidence that people use shifts of attention to gather information about a scene when needed, rather than depending on memory (O'Regan, 1992; O'Regan, Deubel, Clark, & Rensink, 2000). Performance in the change detection task may also be a special case, however, since sudden changes to displays are highly unlikely in the real world. Under such unusual conditions, preferences to trust the visual display could supercede reliance on memory (Becker, Pashler, & Anstis, 2000).

The early studies showing that visual scene recognition is excellent (Potter, 1976; Biederman, 1972; Kundel & Nodine, 1975; Loftus, 1972), like the prior eye movement and ‘change blindness’ work, also do not provide an unambiguous assessment of memory. In some early experiments on ‘gist’ memory, for example, semantic cues could have contributed to accurate guessing about what was in the scene without needing to encode scene details (Biederman, 1972). A reluctance, or inability, to encode scene details due to dominance by ‘gist,’ or by a few vivid features, might also have contributed to poor performance in the change detection tasks (Hollingsworth & Henderson, 2000).

In summary, serious questions remain about how information about a visual scene accumulates across separate glances and over time, and what role, if any, memory plays in the guidance of saccadic eye movements. There are three goals of the present study:

The first goal is to examine the capacity and persistence of visual scene memory across saccades and over time. Unlike previous experiments, subjects in the present study were given a memory test without any competing tasks that might cause memory to be underestimated, and without semantic ‘gist’ or practice effects that might cause memory to be overestimated. The stimuli were computer-generated scenes showing rooms with a wide variety of semantically-unrelated objects placed on shelves, tables, or other surfaces. Subjects were shown the room displays for varying durations, and then given a recall test to measuring the capacity of memory for objects in the scene as a function of viewing time.

The second goal is to examine memory persistence. To do this, a subset of displays were re-tested later in the session to look for any ‘savings’, or improvement in performance. If subjects used a temporary short-term memory to rehearse a list of object labels, then little or no improvement would be expected on re-tests administered many trials later, either because of the passage of time, or because of the interference from the intervening displays. An improvement in the number of items
recalled on the second or third viewing, however, would suggest that scene memory persisted over time in a more enduring representation than the typical short-term memory store. (A brief report of performance with re-tested displays is found in Melcher, in press).

The third goal is to describe the eye movements made while looking around the room displays. This is useful for describing scanning strategies and the relationship between which items were fixed and which were remembered. But, more importantly, information about eye movements pertains to the more basic question of how memory was used, not only for the later recall of items, but also for immediate saccadic planning.

2. Method

Subjects viewed a display of 12 objects in a scene, and then were given a memory test. Display durations were 1, 2 or 4 s. Some displays were re-tested on a subsequent trial to look for a longer-term retention. Eye movement patterns were analyzed to determine scanning strategy.

2.1. Subjects

Three naive subjects, BS, SE, and VI, participated in the experiment. All had normal vision without correction.

2.2. Eye movement recording

Two-dimensional movements of the right eye were recorded by a Generation IV SRI Double Purkinje Image Tracker (Crane & Steele, 1978). The left eye was covered and the head was stabilized on a dental biteboard.

The voltage output of the Tracker was fed on-line through a low pass 50 Hz filter to a 12-bit analog to digital converter (ADC). The ADC, controlled by a PC, sampled eye position every 10 ms. The digitized voltages were stored for later analysis.

Tracker noise level was measured with an artificial eye after the tracker had been adjusted so as to have the same first and fourth image reflections as the average subject’s eye. Filtering and sampling rate were the same as those used in the experiment. Noise level, expressed as a standard deviation of position samples, was 0.4° for horizontal and 0.7° for vertical position.

Recordings were made with the tracker’s automatically movable optical stage (auto-stage) and focus-servo disabled. These procedures are necessary with Generation IV Trackers because motion of either the auto-stage or the focus-servo introduces larger artifactual deviations of Tracker output. The focus-servo was used, as needed, only during intertrial intervals to maintain subject alignment. This can be done without introducing artifacts into the recordings or changing the eye position/voltage analog calibration. The auto-stage was permanently disabled because its operation, even during intertrial intervals, changed the eye position/voltage analog calibration.

2.3. Stimulus

Each stimulus display was constructed using Open Inventor (version 2.0, Silicon Graphics). They consisted of objects seen inside virtual ‘rooms’ with three walls, a ceiling, and a floor (see Fig. 1).

The walls were either colored, as if painted, or covered with a texture that appeared to be wallpaper or wood. There were 14 possible wall coverings (10 colored walls, four textured images). In addition, the floor was covered in one of 10 ‘carpet’ texture patterns. In all, there were 140 different possible wall-floor combinations.

The room, unless otherwise indicated, also contained two to five pieces of 3D ‘furniture’ of the following types: round table, rectangular table, desk, bookshelf, wall shelf, dresser/cabinet, or wooden chest. The placement of furniture was constrained by the requirement that no item could spatially overlap or occlude another item. After placing each item, the number of potential locations for the test objects was calculated based on the width and number of support surfaces. A desk, for example, had only one support surface (the top), while a bookshelf would have one support surface for each shelf. Furniture was placed until 12 or more potential object locations were created. Any arrangement of furniture that did not allow for the placement of 12 objects was discarded.

The objects shown on each trial were randomly chosen from a set of 103 items and placed in potential locations. All objects could be given an identity label such as mug, cat, hammer, hydrant, boot, bucket, plant, or statue. Objects were either taken from public domain web sites or created using Open Inventor, and then scaled and modified to fit within the size and color parameters of the study (see Fig. 1 for examples of objects).

Some objects were more easily labeled than others. To examine the possible role of recognition and verbal labeling on both memory and saccades, the object stimulus set was divided into two groups. Subjects were shown examples from one of the groups before running any trials and instructed to name each object in the set.

The trials in the initial testing sessions were divided so that all objects within each display were from the same set. After the initial sessions, subjects had seen all of the objects at least once.

After the objects locations were chosen for each trial, the 2D screen coordinates were calculated based on
camera position. The 2D location of each object was examined to ensure that none fell within 1° of the fixation cross because pilot studies had suggested that the presence of an object immediately at fixation could be distracting.

The viewing angle of the ‘camera’ within Open Inventor varied horizontally and vertically by about ±5% of the width of the ‘room’ across trials. This gave the impression of viewing the room from different positions and increased the subjective difference between the displays.

Fig. 1. Examples of scenes used in the experiments, along with eye movement patterns.
The entire display subtended 10° by 8°, with each object (depending on location in room and thus distance from the ‘camera’) 1–2° in size. The size of an object was measured as its maximum extent in height, width, or depth. In pilot studies, it was determined that at these sizes, observers could recognize objects in the corner of the room while maintaining fixation on the central cross. Thus, any scanning eye movements were related to the needs of memory, not visual acuity.

Luminance and color were selected so that all objects were easily seen against the backgrounds. Background wall luminance ranged from 5 to 20 cd/m². Objects were either very dark (< 2 cd/m²) or else brighter than the background (> 25 cd/m²). Luminance values were measured using a Minolta ChromaMeter CS-100. Care was taken to ensure that the background and the objects had different chromaticity. Objects included a wide variety of colors and textures.

Each display of background and objects was constructed off-line and saved to a file for later use.

2.4. Procedure

At the beginning of each trial, the subject fixated a central crosshair, and began the trial when ready by means of a button press. The stimulus scene was displayed for a fixed duration within each block of 20–50 trials. Unless otherwise indicated, all blocks contained a single task or response condition, which was either free recall, two alternative forced-choice recognition, cued-recall, or search. Results from the free recall task are reported in this paper, while other measures will be discussed in later reports. Trial durations were 1, 2 and 4 s. These durations were chosen based on preliminary studies that showed an improvement in recall after 1–2 s when subjects were permitted to make saccades, rather than required to maintain fixation.

In the free recall condition, subjects were instructed to begin naming objects from the scene as soon as the stimulus frame was over. Responses were recorded on audio tape for later analysis. Average inter-trial interval, including display and response time, ranged between 30 s for 1-s display durations up to about 40 s for 4 s displays.

On some trials, a stimulus viewed previously was re-tested. Re-tests occurred 1–8 trials later, so that on average, there were three intervening trials between repeats.

Calibration factors were determined in separate sessions and confirmed by additional calibration trials within each session. For these additional trials, the calibration display was a 3 × 3 grid of white balls in front of a gray background room. Subjects were instructed to look from ball to ball during the duration of the trial, in any order they wished. Analysis of these calibration trials was used to determine correction factors applied to eye movements recorded within the same session. These corrections proved to be quite small (within 10% of each other).

2.5. Detection and measurement of saccades

The beginning and end positions of saccades were detected by means of a computer algorithm employing an acceleration criterion. Specifically, we calculated eye velocity for two overlapping 20-ms intervals. The onset time of the second interval was 10 ms later than the onset time of the first. The criterion for detecting the beginning of a saccade was a velocity difference between the samples of 300/s or more. The criterion for saccade termination was more stringent in that two consecutive velocity differences had to be less than 300/s. This more stringent criterion was used to ensure that the overshoot at the end of the saccade would be bypassed. The value of the criterion (300/s) was determined empirically by examining a large sample of analog records of eye position. Saccades as small as the microsaccades that may be observed during maintained fixation (Steinman, Haddad, Skavenski, & Wyman, 1973) could be reliably detected by the algorithm.

The size of each saccade was defined as the distance between the position of the eye at the start of the saccade and the end of the saccade. Eye movement records were also examined manually to verify that the algorithm was detecting all saccades. Occasionally, small adjustments to the criterion were needed to achieve completely accurate detection.

2.6. Number of trials tested

SE was tested in eight sessions, VI in six sessions, and BS in seven sessions of 20–50 trials each.

3. Results

Memory and eye movement performance were analyzed to determine the capacity and duration of scene memory, to examine the use of eye movements in the memory task, and to describe the way that memory was used for guiding saccadic target selection.

3.1. Visual scene memory capacity

Scene memory improved with increasing display duration. The number of objects recalled increased from a mean of three to four items in 1-s trials up to a mean of about five objects in 4-s trials (Fig. 2).
3.2. Memory persistence over time

Scene memory persisted beyond a single trial. Recall performance improved when a previously viewed display was re-tested, such that more items were recalled on the second or third viewing of a display than on previous tests. Fig. 3 shows the number of items recalled as a function of total viewing time. Total viewing time was calculated by multiplying trial duration by the number of times the display had been presented. For example, re-testing the same display twice for 1-s each yields a total viewing time of 2 s. Fig. 3 shows the number of items recalled on the final re-test trial itself, irrespective of previous performance in earlier tests of that display. Repeated displays were always shown for the same duration as the earlier tests.

The number of items recalled increased with total viewing time. The number of items recalled on each re-test of a display was greater, on average, than the number that was remembered the last time the scene was displayed. Subject BS, for example, recalled an average of 3.8 items the first time she saw a display. The fourth time she saw the display the mean number of items recalled was five. Despite the many intervening trials, memory continued to persist and to accumulate.

Fig. 4 compares performance on re-tests with performance on continuous presentations, keeping total viewing time constant. To equate performance in the two different conditions, the abscissa again shows total viewing time (trial duration × number of presentations), which, for a single continuous presentation, is equal to the display duration. Performance as a function of total viewing time was nearly identical in continuous and repeated displays. The same results were obtained with three other subjects, whose eye movements were not recorded (see also Melcher, in press, Melcher, 2001b). Thus, not only did memory persist over time, it summed over repeated trials, regardless of their duration, as if there were no intervening trials (performance on continuous and re-test trials of the same total viewing time was not significantly different for any subject: t-test, P > 0.05). The consistent rate of build-up indicates that memory persisted and summed across trials as if the display had never left the subjects’ view. There was no improvement in recall, however, when a scene was re-tested on a subsequent day (Melcher, in press, Melcher, 2001b), suggesting the memory for the objects was not consolidated into a long-term store.

3.3. Properties of the eye movement patterns

In this section, the properties of the eye movement pattern are analyzed to examine how saccades were used during the memory task.

3.3.1. Saccade size

Subjects often made small eye movements within the vicinity of the same object (14–28% of total saccades),
Fig. 3. Number of items recalled as a function of total viewing time for 1-s re-test trials (BS: triangles, SE: squares, VI: circles). Total viewing time for these trials is equal to the number of times the display has been shown multiplied by 1 s. Data for 1 s trials indicates trials which were later re-tested.

Fig. 4. Number of items recalled for continuous presentation (triangles), 1-s re-test trials (squares), and 2-s re-test trials (circles). Error bars show 1 SE.
Fig. 5. Mean vector saccade size and mean inter-saccadic interval as a function of display duration (BS: triangles, SE: squares, VI: circles). Trials were blocked by duration.

3.3.2. Inter-saccadic intervals

Inter-saccadic intervals increased with display duration (Fig. 5b). This increase suggests a change in global strategy between blocks of different display durations. Interestingly, the increase in ISI for 4-s trials was not observed until after the first 10 trials at this duration. After the first 10 trials, subjects increased mean inter-saccadic interval by about 20 ms (BS: from 373 to 391; SE: from 319 to 336; VI: from 321 to 340), which suggests that a global change in strategy was invoked, perhaps in an attempt to improve performance or to adjust saccade rate to a more comfortable level.

3.3.3. Fixation positions with respect to objects

Did subjects look at individual objects, at groups of objects, or at empty space? For all three subjects a majority of fixations were within 1° of the center of an object (BS: 84.9%; SE: 86.1%; VI: 85.2%). As shown in Fig. 1, subjects tended to look at objects rather than the background. Fig. 1 also shows that some fixations were targeted to a group of objects, since the saccade landed in the middle (within 1.5°) of two or more adjacent objects (BS: 6.7%; SE: 6.7%; VI: 6.9%). The remaining fixations were not within 1.5° of any object.

3.3.4. Fixation and recall performance

It is reasonable to assume that recall would be better for fixated items. In reality, however, the situation proved to be more complicated. The proportion of recalled items that were fixated increased as a function of display duration (Fig. 6). At longer durations, there was a strong link between where the subject looked and what they remembered. At shorter durations, subjects tended instead to recall many non-fixated items, since fewer objects were fixated on these trials.

Finding that items that not directly fixated were, nevertheless, recalled, is consistent with pilot studies showing that eye movements were not important with display durations under 2 s. For these short durations, performance when maintaining fixation on a central cross was equal to performance when scanning freely, as is consistent with prior reports using other visual tasks (He & Kowler, 1992; Kowler & Steinman, 1979; Schlingensiepen, Campbell, Legge, & Walker, 1986).
3.4. Immediate memory and the guidance of eye movements

3.4.1. Immediate memory for objects in the current display

In addition to serving the purpose of building scene memory, information from individual fixations also could be used to plan subsequent eye movements. There could be several potential strategies for using information from prior fixations to plan new saccades. One strategy would be to look at a new object with each saccade. Another would be to revisit the same object several times in an effort to encode the item into memory. Both of these strategies require that the subject keep track of which objects were already fixated. An alternative possibility, however, is to select items randomly without taking into account which objects had been previously seen.

Fig. 7 shows predictions based on these different scanning strategies. The figure shows the cumulative number of new objects fixated as a function of ordinal number of the saccade in the sequence. The straight line shows what would happen if the subject looked at a new item on each saccade, until all objects were seen. The curved line below it shows the number of items that would be fixated if the subject randomly selected an object on each fixation, with replacement (obviously, fewer new items are seen with this strategy) (see Schlengensiepen et al., 1986, for a similar account of saccades during a search task). Data falling above the random selection curve indicates a bias to look at new items and ignore old ones, while data below the random curve indicates a bias to revisit old items.

The data for all three subjects in 4-s trials fell below the predictions of the random selection model (Fig. 7). This tendency to fixate the same object more than once might be explained by a strategy of remembering which items had been seen, and then choosing to revisit them. Although this strategy seems plausible, the assumptions inherent in the random selection model are unrealistic. True random selection implies that subjects would be equally likely to jump all the way across the display as to look at an adjacent object. In reality, however, subjects tended to avoid making large eye movements, and, instead, tended to fixate nearby items, or to make multiple fixations within the same object. The distribution of saccade sizes shown in Fig. 8 confirms this tendency to avoid large saccades.

To take the preference to look toward nearby objects into account, the random sampling model was modified to include a preferential weighting for nearby items. The weighting of each object was determined by its distance from present fixation. The exact values of the weights were chosen using the observed frequencies of saccades of different sizes (Fig. 8).

The weights were then used to obtain a Random Selection with Distance Weighting (RSDW) model. The predictions of this model were derived as follows: From the central fixation position, a target was selected randomly using the weights given to each object based on

![Fig. 6. Proportion of recalled items that were fixated during the trial (BS: circles, SE: squares, VI: triangles).](image-url)
Fig. 7. Cumulative number of unique objects fixated so far in the eye movement sequence as a function of ordinal saccade number. The straight line shows predicted performance if subjects fixated a new object with each saccade. The curved line shows predicted performance for simple random selection with replacement. Data is shown for the three subjects (BS: triangles, SE: circles, VI: squares).

Fig. 8. Distribution of vector saccade sizes for subject SE in 4-s trials. The vertical line shows average distance between objects in that session. The curve shows the exponential function $a \times e^{-bx}$ fit to the distribution, where $a = 2.8911$ and $b = 0.0235498$. 
its distance from the current fixation. After the target of the saccade was selected, the next fixation position was determined. Rather than assume that subjects would fixate the exact center of the object, the simulation was made more realistic by choosing the saccadic error randomly from the obtained distribution of errors for the particular trial duration. Then, starting from this new fixation position, the next target was again selected randomly with the same distance-weighting function. This process was completed until the number of simulated fixations per trial matched the subject's average. A total of 400 simulation trials were generated for each subject using a stimulus set of 20 different displays.

The RSDW model provided an excellent account of the number of new items that had been fixated at each ordinal position in the sequence, as shown in Fig. 9. Thus, performance can be explained without assuming that the saccadic system kept track of which items in the display had been fixated.

3.4.2. Memory and eye movement patterns for repeated displays

As noted earlier (Section 3.2), some displays were tested more than once. What strategy was used to look at a scene that had previously been viewed? If subjects kept track of which objects had been fixated, then they might either look at the same objects again, or conversely, look at all new objects. If subjects were sampling randomly, however, they would look at some old objects and some new objects. A preference to look at nearby items would tend to bias saccades towards the same clusters of objects on repeated trials of the same display.

To find out whether random selection could account for performance, the RSDW model was used to simulate performance with repeated presentations of the same display. The fixation locations obtained from the simulation were examined to determine the proportion of 'old' objects that were fixated during the re-testing of a display. Fig. 10 shows the data along with the model predictions. With one exception (SE's 4 s trials), the random selection model fits performance quite well. The ability of the RSDW model to account for target selection in re-test trials again suggests that target selection was not influenced by memory for which objects had been fixated earlier.

4. Discussion

Visual scene memory accumulated and persisted across saccades and over time. This memory, evident from the accurate reports of the objects in the scenes, was apparently not used in the selection of saccadic targets. This discussion will focus on the implications of these different task-based applications of scene memory.
4.1. Visual memory: keeping the scene ‘in mind’

One unexpected finding was that the number of items recalled after several re-tests of the same display, separated by many intervening displays, was nearly identical to performance in continuous presentations of the same total duration (Fig. 4). This property of memory summation has previously been reported only for inter-stimulus intervals of a few 100 ms (Loftus, Duncan, & Gehrig, 1992). In the current study, the mean time duration between re-test trials was about 2 min. The amount of time between re-tests exceeds the capacity of short term or working memory, according to current models.

The scene memory representation appears to be visuo-spatial in nature, rather than consisting of a verbal list of item names. Several arguments support this claim. First, because subjects never knew whether a given scene would be re-tested, the average number of intervening items that would have to be stored between re-tests was large, greatly in excess of short-term memory capacity for lists of names. Specifically, there were up to eight displays of 12 objects each, with an average of 10–15 (depending on trial duration) objects recalled, between each re-test. Subjects never knew which of these objects, if any, would be re-tested. This uncertainty, as well as the continual presentation of new displays and new objects to be remembered, precludes using a rehearsal strategy to aid recall. Secondly, separate experiments showed that recall for the objects in visual scenes differed from memory for words. In a separate study, the pictorial objects in the scene were replaced with word labels, such as ‘HYDRANT.’ If the memory accumulation were due to verbal codes, then performance in re-tests should be the same for objects as for words. Re-tests using the words resulted in little or no accumulation across separate re-test trials, despite the same rate of initial recall for words and objects immediately after the trial (Melcher, in press, Melcher, 2001b). Finally, recall for objects was linked to memory for the visual background. On some trials, previously-viewed backgrounds were re-tested with new objects. If subjects were only remembering a list of objects, then backgrounds should not be important. Instead, the old backgrounds interfered with recall of the new objects, suggesting that the objects and their background are linked into a larger representation of the whole scene (Melcher, in press, Melcher, 2001b). Interestingly, when words were tested in place of the objects, there was no influence of repeating previously-viewed backgrounds on recall. These findings strongly suggest that a visual representation, and not a verbal list, underlies the accumulation of scene memory.

The build-up of scene memory found in this experiment may reflect a process that ‘keeps in mind’ recent visual information. Exposure to a novel scene presents
the problem of having to decide what to pay attention to and what to commit to memory, since it is impossible to memorize all aspects of the environment with a single glance, nor does it seem efficient to try to do so. It is useful to be able to keep in mind the location and identity of important objects in the immediate environment, particularly in tasks requiring localization and manipulation of objects. When engaged in visuomotor tasks lasting anywhere from several seconds to minutes, people may initially need to search the scene for important objects, but soon become adept at locating what they need without extraneous searching eye movements (Epelboim et al., 1995; Hayhoe, Bensinger, & Ballard, 1998; Land, 2000). The visual memory that facilitates localization of previously-viewed objects may be the same visual memory we found to be evident in the build-up of recall performance with repeated views. The build-up was so effective that even exposure to several intervening displays did not deter the improvement of memory.

It is not likely, however, that the memory was consolidated into true long-term memory. The accumulation of memory with re-test trials did not extend across separate days (Melcher, in press, Melcher, 2001b), and so memory may best be termed ‘medium-term’. Studies of learning in a variety of domains have supported the existence of such ‘medium-term’ storage (McClelland, McNaughton, & O’Reilly, 1986; McGaugh, 2000; Rossi-George & Rovee-Collier, 1999).

4.2. Implications for ‘change blindness’

The excellent accumulation of memory demonstrated by the present results seems at first to contradict prior reports of poor memory representation in studies of ‘change blindness’ (see Section 1). But, encoding limits, rather than unavoidable limits on memory, might have been the limiting factor in the change detection task (Wright, Green, & Baker, 2000). Most change detection tasks have used naturalistic pictures in which scene semantics can lead subjects to ignore, rather than encode, visual details that do not affect the scene’s overall gist (Friedman, 1979; Hollingsworth & Henderson, 2000). In studies of displays without scene semantics, change detection performance depends on the number of objects present, with subjects detecting changes in the same number of items that could report in the classic ‘whole report’ paradigm (Pashler, 1988). When changes occur to an item that has already been fixated, or is the target of the next saccade, the change is usually noticed (Hollingsworth & Henderson, 2000; Hollingsworth et al., in press). Thus, changes made to attended items are typically detected accurately.

4.3. Memory and saccadic target selection

While people can remember what objects are in a scene, subjects in this task did not appear to use this memory to select saccadic targets. Eye movement patterns while scanning the display, as well as when looking at a previously seen display, could be predicted by a random selection model, with items weighted by their distance from the current fixation position. The same excellent memory demonstrated by the accurate recall of items did not appear to be used for the immediate decisions about where to look.

Why was memory not used in saccadic target selection? A plausible explanation is that retrieval is computationally expensive. Checking memory is effortful, and becomes more computationally expensive as the number of items in short-term memory increases (Sternberg, 1969). Eye movement selection strategies in other tasks have also been shown to stress simplicity (Hooge & Erkelens, 1998; Kowler, 2000) and a preference to use eye movements to re-visit previously seen objects (Ballard et al., 1995; Epelboim & Suppes, 2001).

5. Conclusion

The psychophysical reports of memory we obtained suggest that the capacity to remember the detailed contents of a visual scene is excellent, without the aid of any semantic cues. The eye movements, on the other hand, were guided by a memory-less, random selection strategy, suggesting no memory build-up at all. This discrepancy shows that caution is needed when trying to infer cognitive states from measures of eye movements (Viviani, 1990). Had either measure—psychophysical or oculomotor—been obtained by itself, a distorted view of memory and its uses would have resulted. Instead, examining the results obtained from both sets of measures leads to a different conclusion.

Visual memory is excellent both in capacity and persistence, but it is not used to guide the immediate selection of saccadic targets. The simpler, random selection strategy worked quite well. With such a strategy, it is possible to look at many items in a short period of time, remembering most, without having to re-check the status of memory before each saccade. This undemanding, retrieval-free strategy frees cognitive resources for other, more pressing, jobs such as encoding the items that were seen into a durable representation that lasts several minutes. When cognitive resources are free to focus on building up memory, the capacity and persistence of visual memory that is observed far exceeds current estimates.

6. Uncited reference

Busey, & Senders, 1993; Posner & Cohen, 1984; Schacter & Cooper, 1994; Tulving, 1962

Acknowledgements

The research was supported by grant F49620-99-0080 from the Air Force Office of Scientific Research, Life Sciences Directorate, to E.K and by NIMH grant 5 F31 MH12468-02 to D.M. We thank Professor Sven Dickinson, Jacob Feldman, Ilona Kovacs, Carolyn Rovee-Collier, Randy Gallistel and Zili Liu for helpful comments during the course of this research.

References


