

Accumulation and persistence of memory for natural scenes

David Melcher

Departments of Psychology, Oxford Brookes University,
Oxford, United Kingdom



Although our visual experience of the world is rich and full of detail, visual short-term memory (VSTM) can retain only about four objects at a time. Long-term memory (LTM) for pictures lasts longer but may rely on abstract gist, raising the question of how it is possible to remember details of natural scenes. We studied the accumulation and persistence of memory for pictures shown for 1–20 s. Performance in answering questions about the details of pictures increased linearly as a function of the total time that the scene was viewed. Similar gains in memory were found for items of central and marginal interest. No loss of memory was found for picture detail over a 60-s interval, even when observers performed a VSTM or reading task during the delay. Together these results suggest that our rich phenomenological experience of a detailed scene reflects the maintenance in memory of useful information about previous fixations rather than the limited capacity of VSTM.

Keywords: visual memory, natural scenes, eye movements, working memory

Introduction

The ability to remember the contents of visual displays is critical for a variety of tasks. Traditionally, this process has been characterized by storage of around four items in visual short-term memory (VSTM) followed by either information loss or transfer into long-term memory (LTM) (Baddeley, 1992; Cowan, 2001; Sperling, 1960). A number of behavioral and physiological studies have suggested a fixed capacity for VSTM of about three or four items based on the finding that performance in a change detection task drops for stimulus sets larger than about three objects (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Todd & Marois, 2004; Vogel & Machizawa, 2004). Such findings have led to claims that little or no visual detail is retained across separate glances (Horowitz & Wolfe, 1998; O'Regan & Noe, 2001).

Studies of eye movements and behavior in natural tasks have suggested that information about the identity and location of objects is used to guide eye and body movements efficiently (Land, Mennie, & Rusted, 1999). There is also considerable evidence that LTM for natural scenes is extensive (Standing, 1973) and requires only a brief glance at the picture (Potter, 1976). It has been suggested, however, that LTM is abstract and depends on general scene layout and gist rather than exact visual details (Ballard, Hayhoe, & Pelz, 1995; Friedman, 1979; Irwin & Andrews, 1996; Miller & Gazzaniga, 1998; Rensink, 2000).

The current experiments examined the accumulation and persistence of memory for visual details over a period of seconds. This is a time frame that is often important in natural behavior and requires learning about the location and properties of objects across separate glances. There is support for the idea that useful visual information accumu-

lates across separate glances (for a good review, see Tatler, Gilchrist, & Rusted, 2003). For example, Tatler et al. (2003) found that performance in answering questions about the visual details of a photograph increased over a period of seconds. Likewise, there is evidence of LTM for objects from natural scenes (Friedman, 1979; Henderson & Hollingworth, 2003; Hollingworth & Henderson, 2002; Melcher, 2001; Melcher & Kowler, 2001; Parker, 1978).

The goal of the present study was to examine the buildup and persistence of working memory for different visual attributes in a complex image. Natural scenes contain a myriad of visual information, including color, texture, shape, absolute and relative object location, object identity, overall spatial layout, and scene gist. Recent studies suggest that different visual features may have different rates of memory accumulation and decay (Melcher & Morrone, 2003; Pasternak & Greenlee, 2005; Tatler et al., 2003). Such differences in memory for particular visual features might, in fact, help to explain the conflicting reports in the visual memory literature.

A second, and related, issue is the role of scene-related salience in the guidance of attention and subsequent accumulation of memory for objects in a scene. In studies of change detection, alterations to the scene are noticed more quickly when they occur to an item of “central interest” rather than one that is marginal to the general meaning of the picture (O'Regan, Deubel, Clark, & Rensink, 2000; Rensink, O'Regan, & Clark, 1997). This raises the question of which types of items benefit from extended viewing of a scene. One possibility might be that salient items would show more improvement because they are viewed earlier and more often, increasing the chance that their details would be in memory, whereas marginal items would tend to be ignored. On the other hand, marginal items might show the greatest benefit from longer presentations, compensating for a lack of earlier attention,

whereas central items would lose their initial dominance in memory.

This study involved two experiments. In the first, memory accumulation for the visual details in photographs and drawings was tested. On each trial, participants were shown a picture and then were either given a memory test (Figure 1) or instructed to continue on to the next trial. Critically, some images were not followed by a memory probe after the first presentation but only after being shown a second time. This manipulation served to measure whether memory accumulation continued across separate views or was eliminated by the introduction of a subsequent stimulus that should, in theory, fill the capacity of VSTM. Thus, this experiment tested both the accumulation of memory across glances in a single view and the buildup of memory across views of a picture separated by other trials.

The second experiment examined the persistence of scene memory. Specifically, we introduced a change detection or reading task during a 60-s delay period. We chose a color change VSTM task used frequently in behavioral and physiological investigations of working memory capacity (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). A similar task has been shown to interfere with visual search (Han & Kim, 2004). Importantly, the task combined both item color and location to avoid testing only spatial or object working memory.

Methods

Subjects

Twenty-one subjects took part in the first experiment (memory accumulation), whereas the second experiment involved a more in-depth psychophysical study with nine observers (memory persistence). The first experiment lasted approximately 1 hr, whereas the second experiment required 3 hr. Informed consent was obtained from all participants.

Stimuli

Photographs and artistic renderings of natural scenes (drawings and paintings) were taken from non-copy-written images available to the public domain and edited using Adobe Photoshop (Figure 1, top panel).

Each picture contained from 5 to 20 objects (mean of 11.4). The stimuli were displayed on a 21-in. SONY high-resolution monitor viewed from 60 cm and controlled by MATLAB and the Psychophysics toolbox (Brainard, 1997). Each image was approximately 20 deg of visual angle in height and width.

In the first experiment, the memory probe was made up of three written questions displayed on the screen along with three multiple-choice answers to each question



What colour are the flowers sitting on the two wicker tables?

- (1) White (2) Yellow (3) Red

Where is the tree in the pot?

- (1) Bottom right (2) Bottom left (3) Top right

What is the view in the background?

- (1) Housing estate (2) Lake, hills and houses
(3) Snowy scene



Figure 1. Sample stimulus (top panel), question set (middle panel), and forced choice recognition test (bottom panel) used in the experiments (see [Methods](#)).

(see Figure 1, middle panel). Two questions examined visual details about the color or location of objects (e.g., “What was the color of the cup? (1) white (2) blue (3) green”), and the third question asked about the identity of a particular object (e.g., “What object was sitting on the table? (1) plate (2) glass (3) vase”). Questions about the images were written by a group of four people and each person’s questions were evenly distributed throughout the experiment to control for any effects of difficulty. The full set of questions, without any pictures, was given to two naive observers to determine chance performance. Both observers were unable to guess the correct answer at a rate significantly better than chance. The set of pictures used in each experimental block was counterbalanced across participants.

The objects in each scene were categorized as either of central or marginal interest. The criteria used for this distinction were taken from previous change detection studies (O'Regan et al., 2000; Rensink et al., 1997). Objects were defined as “central” based on their relationship to the overall theme of the picture rather than their location on the screen. Given that observers had 5–20 s to view the picture, there was sufficient time to scan any of the objects of interest independent of their location. The theme of each picture was derived by giving a short description of the picture (such as “a bathroom,” “airplanes parked on a runway,” or “three people waiting for a bus”). Two judges worked together to categorize each object based on these descriptions. In the case in which the distinction between central and marginal for a particular object was not clear, that object was not included in the test. Objects were categorized solely on relationship to the gist of the picture rather than dimensions of low-level visual salience (for a review of the role of visual salience on eye movements, see Tatler, Baddeley, & Gilchrist, 2005).

In the second experiment, the memory test was divided over two different display frames. Shape information was tested by a three-alternative forced choice display with three conceptually identical items, only one of which had been present in the scene (Figure 1, bottom panel). Each choice was presented in grey scale. The remainder of the memory probe was made up of the three written questions.

Procedure

Experiment 1

Each trial commenced with a button press. The stimulus display was presented for a period of 5, 10, or 20 s, followed 1 s later by either the memory test or a fixation point. As described above, some images were not followed by a memory test after the first presentation but only after being shown a second time. This manipulation served to measure whether memory accumulation continued across repeats of the same stimuli (Melcher, 2001). Total viewing time for a repeated stimulus was either 10 s (5 + 5) or 15 s (10 + 5). The latter could either be made up of a 10-s duration trial preceded by a 5-s trial (50%), or the opposite pattern (50%). Retest trials were shown four to six trials after the initial display of that stimulus.

The 5- and 10-s trials were run together in two blocks. Each block contained 10 stimuli shown once (continuous trials) and 10 stimuli shown twice before the test, giving a total of 20 trials with questions. The 20-s duration trials were run in separate blocks of 20 trials. The experiment was self-paced, allowing the participant to type in the answer to each question (by giving the corresponding number to their choice: 1, 2, or 3) before moving to the next display screen.

A number of the subjects in the first experiment performed much worse on the memory tests than the rest of the group. The exclusion of these five participants raised average performance by about 10%, making it comparable

to the levels found in the second experiment. The full set of data, however, was used in the analyses and figures because the outliers could not be excluded based on any criterion other than poorer performance.

Experiment 2

In the second experiment, there were five conditions, each containing 20 trials. The first two conditions consisted of displays shown once for 1 or 10 s. The third block of 20 trials contained a 10-s display of the stimulus, followed by a delay of 10 s and then a 1-s redisplay of the stimulus. The fourth block extended the delay to 60 s. During the delay period, a written paragraph (taken from history textbooks) was presented on the screen and participants were instructed to silently read the paragraph throughout the delay period, repeating if necessary. The reading task was designed to occupy visual, verbal, and conceptual working memory. The scenes used in each condition were counterbalanced across subjects, as was the order in which the conditions were tested.

The fifth condition in the second experiment contained a typical VSTM test (Luck & Vogel, 1997). Five colored squares were briefly (500 ms) presented, followed by a 1500-ms blank delay and a further display of five colored squares (500 ms). On 50% of the trials, one of the five squares changed colors and the task was to respond whether the second set of squares was the same or different. After responding, an additional 5-s delay preceded the second view (1 s) of the picture, followed by the usual memory tests. This condition was identical to the other delay conditions except that the reading test was replaced by the VSTM task. The duration of the delay was 12 s plus time for the participant to respond. Performance on this task ranged from 75% to 90%, as would be expected from a VSTM capacity of four or five items.

Results

Performance improved for longer views, $F(1,647) = 8.302$, $MSE = 0.363$, $p < .001$, improving as duration increased from 5 to 10, and then 20 s (Figure 2, solid squares).

This memory accumulation also occurred when a stimulus was shown once without giving the test and then repeated later in a subsequent trial. As shown by the open circles in Figure 2, performance in repeat trials was predicted by the total viewing time (e.g., 5 s + 5 s = 10 s) rather than the duration of each individual presentation of the stimulus. Memory for scene details significantly improved for repeat trials compared to trials of the same duration containing scenes shown only once, $F(1,756) = 6.755$, $MSE = 0.240$, $p < .02$. Overall, performance

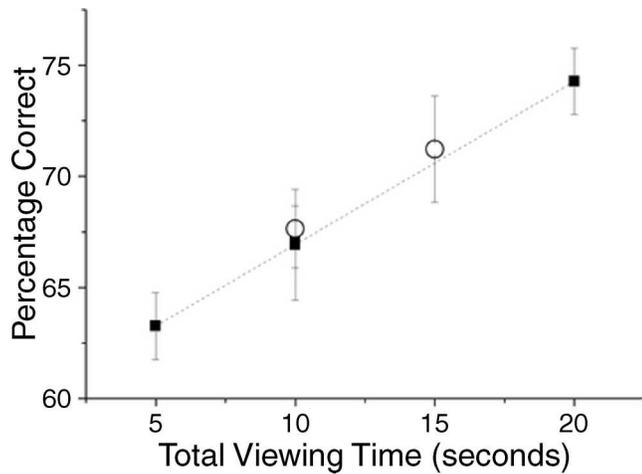


Figure 2. Average performance in [Experiment 1](#) for all question types as a function of total viewing time. Filled symbols show performance on single, continuous trials of a given duration (5, 10, or 20 s), whereas open symbols show percentage correct as a function of total viewing time for a repeated stimulus (see [Methods](#)). The error bars show between-subject standard error of the mean.

improved linearly over time, consistent with our previous studies of free recall for briefly presented scenes (0.25–4 s; Melcher, 2001; Melcher & Kowler, 2001).

Performance was analyzed separately for each of the different question types, with data divided into trials containing items of central and marginal interest (for more details, see [Methods](#)). Overall, performance was superior for questions about items categorized as of central interest ([Figure 3](#), filled circles), maintaining a consistent advantage of about 15% compared to items of marginal interest ([Figure 3](#), filled squares). The same trend was found for repeat trials (open circles and squares).

Performance improved as a function of total viewing time for all three types of question ([Figure 4](#)). The accumulation of memory was similar for each attribute, with location showing the greatest gain. Given that there were fewer trials per condition with the data divided by question type, the results were less consistent and linear in nature. For color, there was little or no improvement as the viewing time increased from 5 to 10 s. This is consistent with the results of Tatler et al. (2003), who also reported a plateau in picture color memory from about 4 to 10 s. Interestingly, observers in our study then continued to improve as stimulus duration increased to 20 s.

Given that the stimulus pictures contained an average of 11.4 objects that could potentially be the topic of a question, it can be estimated that the number of objects remembered after 5 s was around 5.1, whereas approximately 7.0 objects were remembered after 20 s ($11.4 \times [\%correct - 33.33/66.67]$). This estimate is based on a number of assumptions. The first is that observers would perform at chance (33% correct) if they remembered nothing and not miss a single

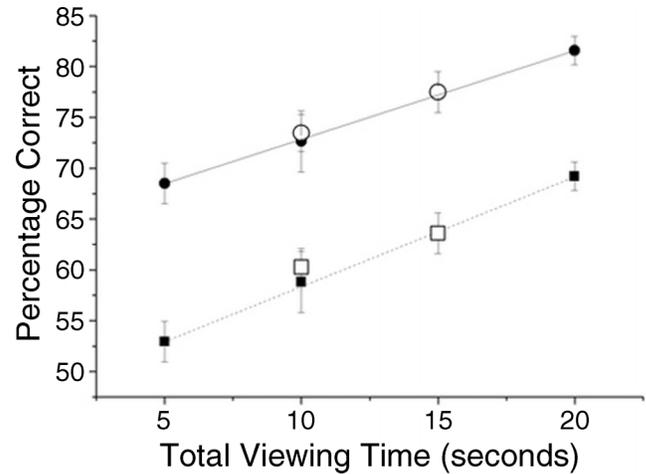


Figure 3. Performance in [Experiment 1](#), divided into questions about items of central interest and items of marginal interest. Notation is the same as in [Figure 2](#).

question (100% correct) if they could remember at least 11.4 objects. In addition, our estimate makes the tenuous assumption that a complete memory for the identity, location, and color of the object was always stored together. It is plausible that observers might correctly answer a question about the identity of an object yet incorrectly report its location or color, in which case the number of items about which some information could be remembered might be far greater.

The results of [Experiment 2](#) provided further evidence that information about the scene remained in memory across separate views. Performance after only 1 s was poor, but above chance ([Figure 5](#)). Adding a 10-s preview before the 1-s view led to a dramatic improvement of around 25%, equaling or surpassing performance after a single 10-s view. In other words, visual memory did not start from scratch, but it started where it had left off.

There was no influence of adding a delay between the 10-s view and the test ([Figure 5](#)). Performance was unaffected by the addition of a 10- or 60-s reading task during the delay, showing that semantic working memory did not play an important role in the persistence of scene memory ([Figure 5](#)).

The VSTM task had no influence on picture memory ([Figure 5](#)). To ensure that any small effect of the VSTM task on a particular type of memory (such as color) was not masked by the combination of data for four separate attributes, we examined whether the VSTM test influenced performance for either color or location, which were the two attributes tested in the VSTM task. The addition of the VSTM task had no measurable influence on average performance for either color or location questions compared to other 10-s trials with no delay or with reading delays of 60 s ([Figure 6](#)). Specific shape information required for the three-alternative forced choice recognition task was also unaffected by the VSTM task ([Figure 6](#)).

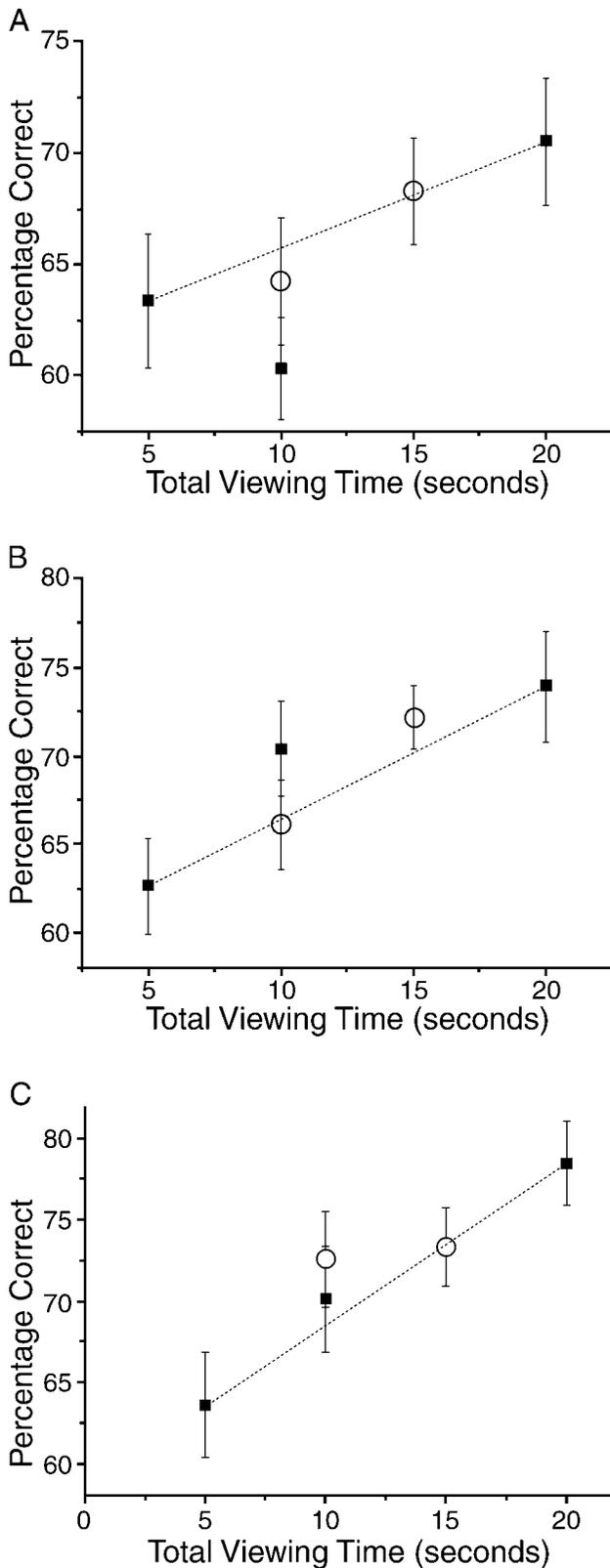


Figure 4. Performance in Experiment 1, divided into different question types. Notation is the same as in Figure 2.

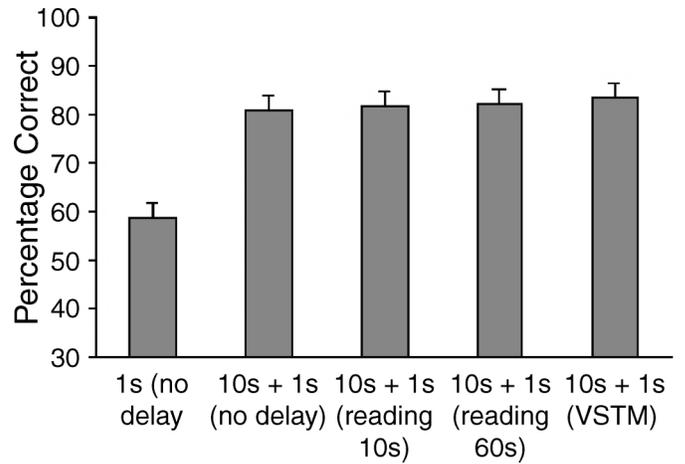


Figure 5. Average performance in Experiment 2 for all question types for the five experimental conditions. The five conditions, from left to right, indicate 1-s trials, 10-s trials, 10-s delay (10-s stimulus, then 10 s of reading text followed by a 1-s stimulus presentation), 60-s delay (as previous, except with a 60-s reading period), and VSTM task trials (with a change detection task replacing the reading delay). Error bars indicate the between-subject standard error of the mean.

It is interesting to note that the three-alternative forced recognition test led to high levels of performance in all of the 10-s trial conditions (Figure 6, rightmost set of bars). Correct performance on the recognition test required that observers remember visual details of object shape because all three choices were similar examples of the same object type (teacup, table lamp, etc.).

Previous studies of the accumulation of shape information have produced inconsistent findings, with no correlation reported between fixation duration and recognition memory for object shape (Hollingworth & Henderson, 2002; Tatler, Gilchrist, & Land, 2005). Performance in this experiment, however, improved as a function of scene viewing time, increasing from 57.2% (1-s trials) to well over 80% (10-s trials), suggesting that information about the shape of

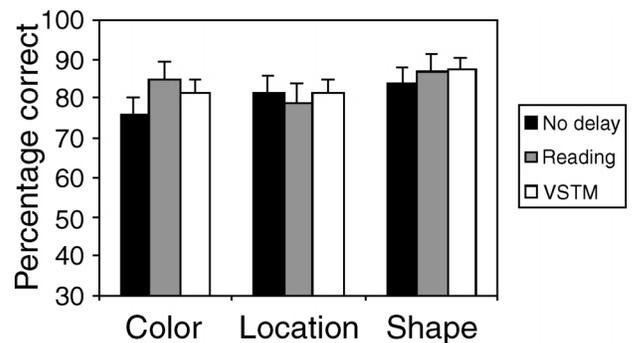


Figure 6. Average percentage correct for color, location, and shape questions in Experiment 2.

multiple objects in a scene accumulates across separate views.

Performance was better overall in the second experiment, with an estimate of about 4.2 objects remembered after 1 s and 8.3 objects after 10 s. The estimated number of items remembered after 1 s is similar to the number of items recalled freely after seeing a scene (Melcher, 2001; Melcher & Kowler, 2001), whereas the number of objects recalled after 10 s is larger than the estimate of 7.4 objects reported in a recent study of object recognition in self-paced trials (Liu & Jiang, 2005).

Discussion

Overall, the current results show that memory for the visual detail of natural scenes accumulates over time and across separate glances. Memory for the visual details of objects in a scene showed a consistent, linear increase over time. It is important to note that this linear accumulation was the same after viewing a scene once, in a single long trial, or after viewing that scene twice for the same total duration. In the latter case, the two views of the trial were separated by an average of five other scenes (Experiment 1) or by a 1-min reading or VSTM task (Experiment 2). The fact that memory accumulation continues across separate trials is contrary to what would have been expected if scene memory relied on a limited VSTM combined with LTM for abstract gist (e.g., Irwin & Andrews, 1996; Rensink, 2000; Todd & Marois, 2004).

Memory for items judged as “central” to the overall meaning of the picture was better than that found for items of marginal interest. This is consistent with findings that changes to these items are noticed more easily (O’Regan et al., 2000; Rensink et al., 1997). The rate of accumulation of memory across the 20-s period, however, was similar for central and marginal interest items, which did not support the hypothesis that memory representation might be built initially with central items and later with less important details. One possibility is that items were fixated quasi-randomly based on low-level attributes (Araujo, Kowler, & Pavel, 2001; Hooge & Erkelens, 1998; Itti & Koch, 2000; Melcher & Kowler, 2001; Parkhurst, Law, & Niebur, 2002) rather than according to the saliency criteria used by our judges. Objects that are not central to the meaning of the scene may still be fixated, although change detection tends to be poorer for these items even when they are fixated (O’Regan et al., 2000; Rensink et al., 1997). Without recording eye movements, however, it is not clear whether the effect of scene-related salience on memory was due to which items were fixated or on how well items were remembered (for a discussion of low-level influences on fixation selection, see Tatler, Baddeley, et al., 2005).

The present results reflect the fact that observers in the current experiments were actively engaged in learning about

the scene, hence their attention was likely drawn to all the items as potential targets for a memory test. In real life, attention may be given primarily to task-relevant items (Land et al., 1999), with ignored objects failing to be retained in memory. Even for the items judged as most salient and viewed for 10 s, observers were able to answer correctly only about 80–85% of the questions. This is a reminder of the contradictory nature of scene memory: it is both abstract and detailed (Gould, 1976; Tatler et al., 2003). While scene memory accumulation is not analogous to stitching together separate snapshots in an internal memory, it does allow for task-relevant details to be accumulated across glances and used to guide our cognition and actions.

Conclusions

The current results are in sharp contrast to suggestions that our visual memory representation is relatively sparse (e.g., Irwin & Andrews, 1996; Rensink, 2000; Todd & Marois, 2004) but add support to the claim that scene memory builds up across separate glances and over a period of minutes (Hollingworth, 2005; Hollingworth & Hollingworth, 2004; Melcher, 2001; Melcher & Kowler, 2001; Tatler et al., 2003). Our initial studies with verbal recall (Melcher, 2001; Melcher & Kowler, 2001) have recently been criticized as failing to eliminate the role of verbal and/or conceptual representations, despite the inclusion of a control study showing that verbal responses to words did not accumulate in the same fashion (Hollingworth, 2003). The present study, which directly probed memory for visual details, may succeed in placating these criticisms. The current paper also provides the opportunity to compare our methods and conclusions with the series of papers by Hollingworth and Henderson on change detection (Hollingworth, 2003, 2005; Henderson & Hollingworth, 2003; Hollingworth & Henderson, 2002; Hollingworth & Hollingworth, 2004; Hollingworth, Williams, & Henderson, 2001). Most importantly, their studies share the conclusion that scene memory is not limited to VSTM, consistent with our findings and those of earlier studies showing accurate LTM for objects in scenes (Friedman, 1979; Parker, 1978). Nonetheless, the exact nature of scene memory and its underlying mechanisms remain a topic of debate.

In our studies (Melcher, 2001; Melcher & Kowler, 2001), which directly examined buildup of memory over time, we found no evidence for a discontinuity in the rate of accumulation that would indicate a transition from STM to LTM, but we did find evidence for information loss over a 24-hr period. This was surprising because memory is traditionally divided into short-term and long-term stores in cognitive models. This led us to suggest that scene memory might involve a proto-LTM, in which some information available over a period of minutes fails to be consolidated

into the permanent memory for the stimulus. Importantly, the theory of “medium-term” memory does not posit the existence of a new, separate memory store (Pierrot-Deseilligny, Murri, Rivaud-Pechoux, Gaymard, & Ploner, 2002). Medium-term memory describes the information that is available in earlier stages of LTM but not consolidated into permanent memory. The theory of consolidation in LTM is a common idea in studies of learning and neurophysiology (McGaugh, 1966, 2000; Müller & Pilzecher, 1900), going back at least as far as the Roman scholar Quintilian (1891).

Thus, despite agreement over the fact that there is LTM for visual details, there remains a fundamental disagreement over whether scene memory can be cleanly divided into two storage systems. The dual stage theory (Hollingworth, 2005; Hollingworth & Hollingworth, 2004) is based on studies of change detection for computer-generated scenes. In a series of papers, Hollingworth and Henderson showed that change detection for objects was better for objects fixated around the time of the change (Hollingworth, 2003, 2005; Henderson & Hollingworth, 2003; Hollingworth & Henderson, 2002; Hollingworth & Hollingworth, 2004; Hollingworth, Schrock, & Henderson, 2001; Hollingworth, Williams, et al., 2001). This is consistent with other reports (O’Regan et al., 2000; Zelinsky, 2001) but may not be surprising, given that performance in difficult tasks tends to improve in general when we look at and/or pay attention to task-relevant items (Nelson & Loftus, 1980; Kowler, Anderson, Doshier, & Blaser, 1995). Moreover, Henderson and Hollingworth found that change detection was best when the changed target was completely different from the original object, whereas performance was still above chance when replaced by a conceptually similar object or a rotated version of the object. Change detection was also superior for recently fixated items (for similar findings of recency effects with natural scenes, see also Irwin & Zelinsky, 2002; Zelinsky, 2001), supporting the claim that change detection involves short-term memory (Luck & Vogel, 1997; Todd & Marois, 2004; Vogel & Machizawa, 2004). While short-term memory appears to play a role in change detection, performance was still above chance for items viewed tens or hundreds of fixations earlier (Hollingworth, 2003), providing further evidence for long-term retention of object details (Friedman, 1979; Parker, 1978). Given that Hollingworth and Henderson found both recency effects and memory over a period of minutes, they proposed that scene memory involves both STM and LTM.

It is interesting to note that change detection actually showed three distinct levels of performance as the number of items viewed between the target and test period was increased (Hollingworth, 2003). Best performance was for items that had just been fixated (better than 90%), followed by performance for items fixation 4 to 10 objects previously (around 83%), with the lowest level of performance for items tested at the end of the session (around 75% correct). These three levels of performance were statistically differ-

ent, which argues against the dichotomous STM/LTM model but is consistent with our finding that the information persisting beyond VSTM is not necessarily consolidated into permanent LTM. A similar finding is reported in another change detection paper (Hollingworth & Henderson, 2002), in which performance was better for items shown earlier in the same trial compared to items shown in previous trials, despite the fact that both situations involved memory beyond the duration and capacity of short-term memory. Although Henderson and Hollingworth do not give importance to the change in performance within the timeframe of LTM, we have suggested that it reflects an efficient strategy to avoid retaining details in permanent store that are unlikely to be useful in the future (Melcher, 2001; Melcher & Kowler, 2001).

Although memory has traditionally been divided, based on duration, into STM and LTM, others have argued that the term “working memory” better describes the online nature of the information that is readily available while engaged in a particular task (Baddeley, 1992). The terms “visual short-term memory” and “visual working memory” are often used interchangeably (Hollingworth & Henderson, 2002; Luck & Vogel, 1997; Vogel & Machizawa, 2004), but this ignores important theoretical differences such as the assumption that the former is a storage stage between sensory memory and LTM, whereas the latter is the collection of all information readily available in mind, some of which may have been recently retrieved from long-term storage (Baddeley, 1992). Studies of complex tasks such as reading have led to the theory of “long-term working memory” (LT-WM), which includes both working memory and accessible aspects of LTM (Ericsson & Kintsch, 1995). This theory is based on findings that skilled behavior can be interrupted and then later resumed without hindering performance and that working memory capacity can increase significantly as a result of practice and expertise (Ericsson & Kintsch, 1995). Because people are experts at viewing natural scenes, the efficient use of LTM for the gist and layout of the scene might increase the capacity and duration of working memory for visual scenes. It is not clear, however, whether the LT-WM model can fully account for the medium-term nature of memory for objects in scenes.

Another issue for scene memory is the fact that buildup depends on the type of information involved, such as shape, color, location, and identity (Tatler et al., 2003; Tatler, Gilchrist, et al., 2005). This would not be predicted by a simple STM-LTM model that treats all kinds of information as identical. The tasks used by Hollingworth and Henderson, for example, have not directly tested what type of information is critical for change detection. Although they tested change detection for different types of changes (see also O’Regan et al., 2000), they did not control for the low-level, metric differences between the different types of change or the nature of the visual features involved. Both low-level (Zelinsky, 2003) and conceptual factors (O’Regan et al., 2000) have been shown to predict change

detection performance. Because Hollingworth and Henderson did not have a low-level metric to compare the visual impact of the changes, the role of conceptual factors can only be inferred. Thus, for example, the difference between an object and its rotated counterpart across a range of visual features might be greater—or lesser—than the difference between two examples of the same object type. In contrast, psychophysical studies of visual memory have allowed for comparisons across stimulus features, finding high fidelity memory for basic attributes such as orientation, contrast, and motion for periods of 10 s or more (for an excellent review, see Pasternak & Greenlee, 2005). These studies have not, however, looked at differences in memory accumulation for different features.

Perhaps the strongest evidence for the accumulation of basic visual information across glances comes from studies of trans-saccadic memory. We have shown, for example, that temporal integration of subthreshold motion continues across saccades (Melcher & Morrone, 2003) and that visual adaptation aftereffects transfer spatiotopically across separate glances (Melcher, 2005). The pattern of results in those studies suggests that memory accumulation is greatest for information that is task relevant, predictive, and invariant across saccades. In the case of visual aftereffects, for example, local information about object contrast is lost across separate glances, whereas mid-level information (orientation and 2-D shape) is partially maintained and high-level visual information pertaining to object identity completely transfers across saccades (Melcher, 2005). Our trans-saccadic memory findings provide a potential explanation for the differences in change detection found for various types of changes (Hollingworth & Henderson, 2002; Hollingworth, Schrock, et al., 2001; O'Regan et al., 2000; Rensink et al., 1997). Exact viewpoint of the object is not invariant across changes in the position of the observer, and thus this type of information would be least likely to persist in memory, even when observers knew that it was task relevant. Given that scene viewing, outside the laboratory, typically involves self-motion with respect to the objects, it would be inefficient to encode in detail the orientation of each object with respect to each view. On the other hand, visual features related to object identity would be most likely to be remembered, leading to better change detection.

Under natural viewing conditions, scene memory is likely to include a myriad of different types of information from the different senses, including semantic gist, visual–spatial layout, visual detail, nonvisual information, and multimodal object representations (Pierrot-Deseilligny et al., 2002; Romo, Brody, Hernandez, & Lemus, 1999). There is evidence that the scene representation accumulates over time, with gist and general layout represented first and details about specific objects added on subsequent fixations (Melcher, 2001; Melcher & Kowler, 2001; Tatler et al., 2003). At the same time, a comprehensive theory of scene memory might be extended to include information that is

below the threshold of conscious report but that nonetheless influences later behavior (Hayhoe, Bensinger, & Ballard, 1998; Ryan & Cohen, 2004; VanRullen & Koch, 2003). All of these types of information may contribute to our impressive ability to learn across separate glances and to interact with the environment in an efficient way.

Acknowledgment

This project was supported by grants from the British Academy (SG37385) and the Royal Society (2003/R2).

Commercial relationships: none.

Corresponding author: David Melcher.

Email: dmelcher@brookes.ac.uk.

Address: Gipsy Lane, Oxford OX3 0BP, United Kingdom.

References

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111. [[PubMed](#)]
- Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: The costs of choosing the optimal path. *Vision Research*, *41*, 3613–3625. [[PubMed](#)]
- Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. [[PubMed](#)]
- Ballard, D. H., Hayhoe, M. M., & Pelz, I. B. (1995). Memory use during hand-eye coordination. *Journal of Cognitive Neuroscience*, *7*, 66–80.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. [[PubMed](#)]
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral Brain Science*, *24*, 87–114. [[PubMed](#)]
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245. [[PubMed](#)]
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*, 316–355. [[PubMed](#)]
- Gould, J. D. (1976). Looking at pictures. In R. A. Monty & J. W. Senders (Eds.), *Eye movements and psychological processes* (pp. 323–346). Hillsdale, NJ: Erlbaum.

- Han, S. H., & Kim, M. S. (2004). Visual search does not remain efficient when executive working memory is working. *Psychological Science*, *15*, 623–628. [PubMed]
- Hayhoe, M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, *38*, 125–137. [PubMed]
- Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics*, *65*, 58–71. [PubMed]
- Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 388–403. [PubMed]
- Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 396–411. [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., & Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 519–537. [PubMed]
- Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory & Cognition*, *29*, 296–304. [PubMed]
- Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, *8*, 761–768. [PubMed]
- Hooge, I. T., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research*, *38*, 1295–1302. [PubMed]
- Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*, 575–577. [PubMed]
- Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance: XVI. Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.
- Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, *64*, 882–895. [PubMed]
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. [PubMed]
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*, 1897–1916. [PubMed]
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311–1328. [PubMed]
- Liu, K., & Jiang, Y. (2005). Visual working memory for briefly presented scenes. *Journal of Vision*, *5*(7), 650–658, <http://journalofvision.org/5/7/5/>, doi:10.1167/5.7.5. [PubMed] [Article]
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. [PubMed]
- McGaugh, J. L. (1966). Time-dependent processes in memory storage. *Science*, *153*, 1351–1358. [PubMed]
- McGaugh, J. L. (2000). Memory: A century of consolidation. *Science*, *287*, 248–251. [PubMed]
- Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, *412*, 401. [PubMed]
- Melcher, D. (2005). Spatiotopic transfer of visual-form adaptation across saccadic eye movements. *Current Biology*, *15*, 1745–1748. [PubMed]
- Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, *41*, 3597–3611. [PubMed]
- Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience*, *6*, 877–881. [PubMed]
- Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, *36*, 513–520. [PubMed]
- Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis [Experimental contributions to the theory of memory]. Leipzig: J.A. Barth. *Zeitschrift für Psychologie. Ergänzungsband*, *1*, 1–300.
- Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 391–399. [PubMed]
- O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, *7*, 191–212.
- O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral Brain Science*, *24*, 939–973. [PubMed]

- Parker, R. E. (1978). Picture processing during recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 284–293. [PubMed]
- Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107–123. [PubMed]
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6, 97–107. [PubMed]
- Pierrot-Deseilligny, C., Muri, R. M., Rivaud-Pechoux, S., Gaymard, B., & Ploner, C. J. (2002). Cortical control of spatial memory in humans: The visuoculomotor model. *Annals of Neurology*, 52, 10–19. [PubMed]
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509–522. [PubMed]
- Quintilian, M. F. (1C AD/1891) *Institutionis oratoriae liber decimus: A revised text*. Oxford: Clarendon.
- Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, 40, 1469–1487. [PubMed]
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368–373.
- Romo, R., Brody, C. D., Hernandez, A., & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, 399, 470–473. [PubMed]
- Ryan, J. D., & Cohen, N. J. (2004). The nature of change detection and online representation of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 988–1015. [PubMed]
- Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, 74, 29.
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25, 207–222. [PubMed]
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45, 643–659. [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 Section A, Human Experimental Psychology*, 58, 931–960. [PubMed]
- Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, 32, 579–592. [PubMed]
- Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428(6984), 751–754. [PubMed]
- VanRullen, R., & Koch, C. (2003). Competition and selection during visual processing of natural scenes and objects. *Journal of Vision*, 3(1), 75–85, <http://journalofvision.org/3/1/8/>, doi:10.1167/3.1.8. [PubMed] [Article]
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in working memory capacity. *Nature*, 428, 748–751. [PubMed]
- Zelinsky, G. J. (2001). Eye movements during change detection: Implications for search constraints, memory limitations, and scanning strategies. *Perception & Psychophysics*, 63, 209–225. [PubMed]
- Zelinsky, G. J. (2003). Detecting changes between real-world objects using spatiochromatic filters. *Psychonomic Bulletin & Review*, 10(3), 533–555. [PubMed]