

Visual memory for objects in natural scenes: From fixations to object files

Benjamin W. Tatler

University of Sussex, Brighton, UK

Iain D. Gilchrist

University of Bristol, Bristol, UK

Michael F. Land

University of Sussex, Brighton, UK

Object descriptions are extracted and retained across saccades when observers view natural scenes. We investigated whether particular object properties are encoded and the stability of the resulting memories. We tested immediate recall of multiple types of information from real-world scenes and from computer-presented images of the same scenes. The relationship between fixations and properties of object memory was investigated. Position information was encoded and accumulated from multiple fixations. In contrast, identity and colour were encoded but did not require direct fixation and did not accumulate. In the current experiments, participants were unable to recall any information about shape or relative distances between objects. In addition, where information was encoded we found differential patterns of stability. Data from viewing real scenes and images were highly consistent, with stronger effects in the real-world conditions. Our findings imply that object files are not dependent upon the encoding of any particular object property and so are robust to dynamic visual environments.

Oculomotor behaviour places specific constraints on the visual system and on the way that we perceive the world around us. High-resolution vision is confined to the central, foveal, portion of the retinal image, and so to sample a particular location in the scene faithfully we must direct our eyes toward it. Slow photoreceptor response dynamics result in image blur during even relatively slow rotations of the eyes. Consequently, eye movements are accomplished by

Correspondence should be addressed to Benjamin W. Tatler, now at the Department of Psychology, University of Dundee, Dundee, DD1 4HN, UK. Email: b.w.tatler@dundee.ac.uk

We thank Sandy Pollatsek for helpful comments during the preparation of this manuscript. We are also extremely grateful for the insightful and helpful comments by Keith Rayner, Ralph Radach, and an unnamed reviewer on an earlier version of this manuscript. This research was supported by a Biotechnology and Biological Sciences Research Council grant (S15868) to BWT.

rapid jerks, called saccades, to minimize visual disruption (e.g., Dodge, 1900; Lamare, 1892). During saccades vision is also actively suppressed (e.g., Burr, Morrone, & Ross, 1994; Latour, 1962; Matin, 1974). Between saccades stabilizing mechanisms maintain a relatively stable retinal image. This ubiquitous saccade and fixate behaviour means that information is gathered from scenes as a disjointed series of glimpses (for recent reviews of the development of our understanding of oculomotor behaviour, see Rayner, 1998; Tatler & Wade, 2003; Wade, Tatler, & Heller, 2003).

In everyday, but complex, visual environments, it may be useful to have access to information outside the scope of the current fixation. For example, it might be useful to remember details of an object (e.g., its location and identity) that we fixated previously and will require at some later point in the task. It might also be useful to have some information about the wider visual environment so that we can relate the current fixation to its broader context in the scene. Whether such information is retained between fixations and the nature of the retained information have been the focus of much research and debate.

For some time it was thought that the visual system might construct a faithful copy of the external scene (e.g., Marr, 1982), formed by retaining and merging the content of each fixation (e.g., McConkie & Rayner, 1976; Rayner, 1978). However, evidence against this suggestion of transsaccadic integration has been available for some time (e.g., Bridgeman, Hendry, & Stark, 1975; Irwin, 1991, 1993; Irwin, Yantis, & Jonides, 1983; McConkie & Zola, 1979; Pollatsek & Rayner, 1992; Rayner, McConkie, & Zola, 1980; Rayner & Pollatsek, 1983). For example, Bridgeman et al. (1975) found that observers were not able to detect a displacement of a visual scene if the displacement occurred during a saccade. Rayner and Pollatsek (1983) found that observers could not integrate dot patterns presented in two successive fixations. Recent change detection research (since Grimes, 1996) has further strengthened the argument against transsaccadic fusion. Observers are poor at detecting changes of object position, colour, or presence if those changes are timed to coincide with saccades (Grimes, 1996; McConkie & Currie, 1996), blinks (O'Regan, Deubel, Clark, & Rensink, 2000), or brief artificial interruptions (Rensink, O'Regan, & Clark, 1995, 1997, 2000). Given this body of evidence, it appears unlikely that the visual system constructs a faithful internal model of the scene.

While complete veridical information may not be retained from each fixation, a large body of evidence demonstrates that some information is extracted and retained (e.g., Henderson, 1994, 1997; Irwin, 1991, 1993; O'Regan & Lévy-Schoen, 1983; Pollatsek & Rayner, 1992; Rayner et al., 1980). Global properties of visual scenes can be encoded and retained, such as their meaning or "gist" (e.g., Biederman, 1981; Introub, 1980, 1981) and a description of the overall spatial layout (e.g., Gibson, 1979; Hochberg, 1968; Rensink, 2000a; Simons, 1996). Information retention is not confined to global scene properties. Object properties such as identity, shape, and colour have also been found to be retained (e.g., Henderson, 1994; Henderson & Hollingworth, 1999; Henderson & Siefert, 1999; Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002; Melcher, 2001; Tatler, Gilchrist, & Rusted, 2003). These findings demonstrate that some information can survive the current fixation and that we retain a combination of object descriptions and more global scene information.

While it now appears clear that we do retain information about the properties of objects in our visual environment, we are only recently beginning to explore and understand the

nature and characteristics of the retained object descriptions. We can characterize these object descriptions by considering two fundamental issues: (a) information uptake—the way in which object properties are extracted for retention from each fixation, and (b) information stability—whether any encoded information is retained stably or transiently. These two issues are explored in the present experiments and are discussed below.

Information uptake

Information about objects in natural scenes appears to accumulate over several seconds of viewing (Melcher, 2001; Tatler et al., 2003). One possibility is that this accumulation of information may reflect the discontinuous sampling imposed by saccades. A number of studies have investigated the role of fixation position in the formation of scene memories (e.g., Antes & Penland, 1981; Currie, McConkie, Carlson-Radvansky, & Irwin, 2000; De Graef, Christiaens, & d'Ydewalle, 1990; De Graef, Detroy, & d'Ydewalle, 1992; Henderson & Hollingworth, 1998; Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002; McConkie & Currie, 1996; Rayner & Pollatsek, 1992).

Reports of preferential encoding at the current point of fixation (e.g., Henderson, Weeks, & Hollingworth, 1999; Nelson & Loftus, 1980) or the target of the next saccade (e.g., Currie et al., 2000; McConkie & Currie, 1996) demonstrate that fixation plays a crucial role in encoding information from scenes. However, while fixation position influences encoding, this result does not in itself demonstrate that encoded information is accumulated across fixations (see Irwin & Andrews, 1996).

Loftus (1972) found that recognition memory for pictures increased as the number of fixations on the picture increased. Irwin and Zelinsky (2002) found that performance in a partial report of object token information increased with total number of fixations on a scene. Hollingworth and Henderson (2002) demonstrated that performance in change detection tasks was elevated for changes made to objects fixated previously compared to changes made to objects that had not been fixated prior to the change. They also showed that the total time spent fixating the to-be-changed object prior to the change event influenced subsequent change detection; increasing viewing time increased detection.

While it is clear that fixation plays a crucial role in encoding the information that is retained from scenes, we still do not fully understand the way in which information from multiple (not necessarily consecutive) fixations on an object is treated. Current evidence is conflicting, with some researchers arguing that information does not accumulate from multiple fixations, but others arguing that it does. Whether information is accumulated across fixations or not has implications for our understanding of object memory and the role of oculomotor sampling when viewing scenes. For example, if information is accumulated, refixating an object will update the retained description of the object. On the other hand, if information is not accumulated from fixations, the pressure to refixate objects might diminish because these refixations do not add to the object descriptions. Under this framework, returning to an object might serve more as a mechanism to check that the object description is still valid. Whether object properties are encoded and accumulated from multiple fixations is of crucial importance to understanding how information is gathered from scenes.

Information stability

Information about objects encoded at fixation either may be retained for some time or may be transient, fading away over the next few fixations of the scene. In recent accounts of visual representation, there has been some disagreement about the temporal stability of retained information. Rensink (Rensink, 2000a, 2000b) proposed a view of scene perception in which detailed object representations are only available for the currently attended object. Once attention is withdrawn, the representation of an object in visual short-term memory dissolves. Apparent lack of memory for previously attended targets in visual search experiments (e.g., Wolfe, 1996) supports this notion that retention of object descriptions is transient.

A less extreme view of transience is offered by Irwin's *object file theory* (Irwin, 1992a, 1992b; Irwin & Andrews, 1996). According to this theory, object descriptions are maintained after attention is withdrawn from an object, but the capacity of visual short-term memory is limited to only three or four recently attended objects. Consequently, retention of information would be transient, but would endure at least for a few fixations after attention is withdrawn from an object, until replaced by a new object file. Irwin and Zelinsky (2002) found evidence for this pattern of transience using partial report of object token information. Performance showed a clear decrease as the number of fixations between the fixation on the item and the test interval increased, with the most notable decrease occurring over the first three fixations.

In contrast to reports of the transience of information retention, Henderson and colleagues have argued that object descriptions are not transient, but are maintained stably for multiple objects in scenes (Henderson & Hollingworth, 1999; Henderson et al., 1999; Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001). Using a change detection paradigm, Hollingworth and Henderson (2002) found no evidence for temporal instability and suggested that the visual object memory that they described was not transient. Hollingworth et al. (2001) found that participants were able to detect object token changes that occurred during saccades away from the changed object. This result argues against transience hypotheses because immediately prior to a saccade, attention is directed to the target of that saccade (e.g., Deubel & Schneider, 1996; Henderson, Pollatsek, & Rayner, 1989; Hoffman & Subramaniam, 1995; Shepherd, Findlay, & Hockey, 1986). Thus at the time of the change, attention had already been withdrawn from the object and so the retained description of that object should no longer have been available.

These possibilities have distinct potential implications for the functionality of retained object descriptions in the real world. Transient object descriptions can be dynamic, which might be advantageous in situations where the visual environment (or objects within it) is changing. Having stable object descriptions means that less frequent resampling of the visual environment is required in order to maintain them; this reduces the need for resampling with the eyes within a relatively stable visual environment.

The present study

The present experiments examined object information retention by measuring its accumulation from multiple fixations on objects and the subsequent stability of that information.

We tested immediate recall of object properties when observers viewed real-world scenes (Experiment 1) or computer-displayed images of the same scenes (Experiment 2). In both settings, multiple object properties were tested at the end of each trial. This avoids an experimental situation where a single object property can be prioritized during encoding and forces concurrent encoding of object properties. The concurrent encoding of object properties is likely to be more ecologically valid.

Our data allow us to explore the two issues discussed above: how information is acquired from fixations and how stably it is retained. To address the former we compared immediate recall performance to two measures. First we compared performance to the number of times the object tested in the question was fixated during viewing, for five object properties: object presence, shape, colour, position, and relative distances between objects. This approach allowed us to consider whether each of these properties is encoded and whether any encoded information is accumulated during revisits to the object. We chose to characterize information uptake in terms of the number of fixations rather than the time spent fixating an object. This follows the suggestion by Loftus (1972) that it is the number of fixations rather than their durations that affects recognition memory. However, one potential limitation of comparing performance to the number of fixations on an object is that we are effectively treating fixations of different durations as equivalent units for information extraction. Consequently we also analysed our performance data to the total time spent fixating an object during viewing (the sum of the durations of all fixations on a target object). Based on previous studies of information accumulation (e.g., Hollingworth & Henderson, 2002; Tatler et al., 2003), we expect to find that all of our tested information types are encoded and retained, but perhaps showing different patterns of accumulation.

To address the question of stability of encoded information we compared immediate recall performance to the number of intervening fixations between the final fixation on the tested object and the end of the trial. This allowed us to identify any recency effects that would demonstrate transience of encoded object properties and so evaluate current discrepancies in accounts of the stability of object memories (e.g., Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002).

Our experiments allow us to extend current understanding of object memory in two ways. First, concurrent information processing conditions are likely to be more ecologically valid than paradigms where single types of information can be prioritized. Therefore we can evaluate whether current accounts of the extraction of object properties from fixations (e.g., Hollingworth & Henderson, 2002) remain valid under such conditions and are therefore more likely to reflect the operation of the visual system under real-world conditions. Second, we can directly compare and contrast information extraction from the same scenes presented either as computer-displayed images or as real three-dimensional rooms. This comparison allows us to evaluate the ecological validity of computer-based paradigms used in the study of information extraction from scenes and object memory. We can also explore and characterize any differences between these two situations. This comparison is important as most of our current understanding of object and scene memory derives from computer-based paradigms.

EXPERIMENT 1

Method

Participants

A total of 27 participants (mean age 25.6 years, $SD = 6.1$) took part in the main experiment. A total of 16 participants (mean age 24.5 years, $SD = 4.6$) took part in the control experiment. None of the participants in the control experiment took part in the main experiment. All had normal or corrected-to-normal vision.

Procedure

Six temporarily empty offices in the School of Biological Sciences at the University of Sussex were set up to mimic a laboratory, an office, a waiting room, a seminar room, a dining room, and a kitchen (Figure 1). Prior to the experiment, participants were informed that they were going to view a series of rooms each for just a few seconds, and that after each room they would be given a questionnaire asking about what they had just seen. They were told that the questions could be about anything in the room that they had just seen.

Each participant viewed all six of the rooms while wearing an eye tracker (see below). At the start of each trial, participants stood in the doorway with the door closed. They were instructed to keep their eyes closed while the door was opened fully. The experimenter then instructed the participant to open his or her eyes. Participants were allowed to view the room freely and were allowed to move their eyes, head, and trunk as they wished but were not allowed to move their feet during viewing. After approximately 5 seconds participants were instructed to close their eyes and immediately turn around to face away from the room. Post hoc analysis of the video records showed that viewing times were 5.0 s ($SD = 0.4$ s) from opening to closing the eyes for all participants. Immediately upon turning to face away from the room participants were presented with a questionnaire about the room just viewed. The questionnaire presented to participants after viewing the “laboratory” room (Figure 1a) is shown in Figure 2.

Each questionnaire comprised nine questions, testing five categories of information about the room: object presence, object shape, object colour, absolute position, and relative distances between objects. Different target objects were tested in each question. One question tested presence but all other categories of information were tested with two questions each: one in the form of a four-alternative forced choice (4AFC), the other as open questions (e.g., “what colour was the notebook?”). For the 4AFC questions one option was always correct, and the foils were objects of similar size and contextual viability. Foils for presence question were objects that might be expected within the room. Foils for shape and colour were viable alternatives for the type of target object. Foils for position questions were positions occupied by other objects in the room viewed. Relative distance foils were other objects in the viewed room.

Two discriminability issues were not controlled for in our experiments. First, we cannot account for any perceptual differences in target and foil discriminability—for example, some objects may have been perceptually more discriminable than others. Neither can we account for any perceptual discriminability differences between question types—for example, colour and shape information extraction may not be perceptually comparable. Hence we cannot make quantitative comparisons between the different question categories, such as directly comparing rates of accumulation or decay of the different types of information. However, it is still valid and informative to make qualitative comparisons between question categories and to make quantitative comparisons within question categories. Qualitative comparisons include comparing the general profiles of accumulation or decay: For example, perhaps some types of information accumulate whereas others do not, or for some a plateau in performance may be reached within a few fixations, but for others accumulation may continue.

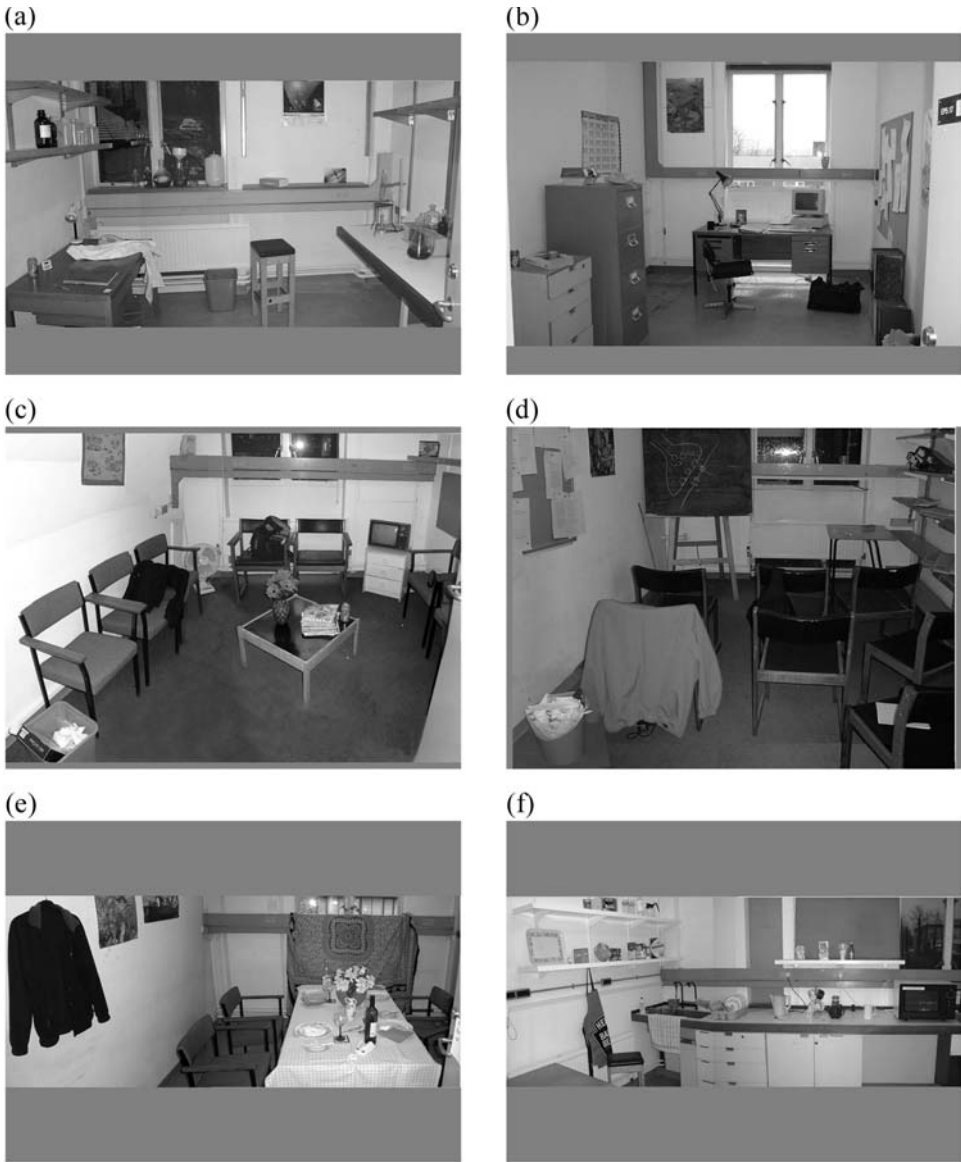


Figure 1. Panoramas of the six rooms viewed in Experiment 1. These panorama were used as the stimuli for Experiment 2.

The nine questions asked after each trial were divided into three blocks, the order of which was varied between rooms, but not between participants. This allowed us to look for any order effects—for example, due to decay of object memory during the course of answering the questionnaire. Predictability of target objects was reduced by having a different target object in each question, by using complex scenes with many potential target objects and by using target objects of varying sizes and positions.

Which **one** of the following was present?

A. Box of matches	B. Phone book	C. Test tube rack	D. Microscope
----------------------	------------------	----------------------	------------------

answer ☐

In the box below, draw the shape of the **tripod**

Name the object that was nearest to the **lab coat**?

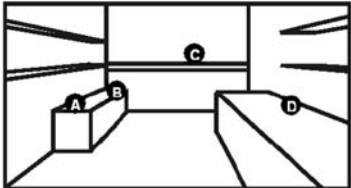
.....

Which **one** of the following is the correct colour for the **pen**?

A. Green	B. Red	C. Blue	D. Black
-------------	-----------	------------	-------------

answer ☐

Which **one** of the following is the correct position for the **phone**?



answer ☐

Which **one** of the following was nearest to the **bin**?

A. Bunsen burner	B. Stool	C. Lamp	D. Drinks can
---------------------	-------------	------------	------------------

answer ☐

What colour was the **notebook**?

.....

Which **one** of the following best matches the shape of the **lamp**?

A. 	B. 	C. 	D. 
---	---	---	---

answer ☐

On the diagram below, mark the correct position for the **Bunsen burner** with a **cross**

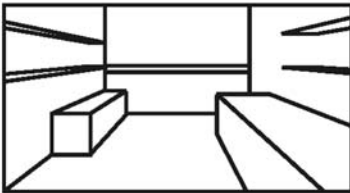


Figure 2. The three sections of the questionnaire given to participants after they viewed the “laboratory” room in Experiment 1 (see Figure 1a). Some small changes in layout have been made to the questionnaire sections to fit them into this figure. This same questionnaire was used in the control experiment and in Experiment 2.

By varying the size and position of target objects we do introduce potential confounds for comparing target objects, but we feel that the costs imposed by this variation are outweighed by the benefits of reducing target predictability and keeping the experimental protocol as naturalistic as possible.

A potential limitation of our protocol is that the use of explicit questioning might limit our ability to test object memories that may well be implicit. To some extent the use of 4AFC questions means that we may be probing some implicit memory, but we have no measure of the extent to which reports were based on explicit or implicit memory. By interrupting the process of information accumulation at different points and in an unpredictable way (after varying and uncontrolled numbers of fixations of each item), responses by participants reflect a translation of the state of the object memory at the time of interruption. Hence reports are a translation of what has been accumulated up to the point of interruption and can therefore be used to infer details of the build-up of information about viewed objects.

Controls

A control experiment was conducted to account for “guess-ability” of questions based on previous experience of similar types of room (see Tatler et al., 2003). Control participants did not view the rooms, but completed a questionnaire comprising six sections, corresponding to the six rooms in the main experiment. In each section, the identity of the (unseen) “scene” was stated and followed by the nine questions about the scene. The nine questions were exactly the same as those asked in the main experiment for each particular “scene”. In this way control participants responded to exactly the same set of questions as did participants in the main experiment. However, control participants only had available past experience and expectation for answering the questions.

Eye movement recording

Eye movements of participants were recorded using the lightweight, portable Land eye tracker used in previous studies of eye movements in everyday activities (e.g., Land, 1993; Land & Lee, 1994; Land, Mennie, & Rusted, 1999). This eye tracker samples at 50 Hz, and with careful calibration and off-line analysis its accuracy is at worst $\pm 1^\circ$.

Trials were excluded from the eye movement analyses if the eye tracker set-up was poor. Poor set-up occurred if there was an unclear view of the eye in its socket, if the part-silvered mirror was positioned such that the positions fixated in the scene were off-screen, or if the calibration could not allow determination of fixation position to within $\pm 1^\circ$. As a result of these criteria 13 trials were excluded from the analyses of eye position data. Within trials, data for individual questions were excluded if participants did not answer the question on the questionnaire. As a result, five individual questions (from 4 different participants) were excluded from analyses. After all exclusions had been made from the eye movement data, a total of 1,237 responses from 149 trials remained.

Analysis

Analysis was carried out off-line after completion of the experiment. Gaze position was compared to the position of objects tested in the questionnaires. An object was judged to have been fixated if the centre of gaze fell within 1.5° of it; this definition follows the angular suggestion for foveal size in a naturalistic task found by Johansson, Westling, Backstrom, and Flanagan (2001).

Saccade detection was manual, from the video sequence (for details, see Land, 1993; Land & Lee, 1994). The minimum detectable saccade size was $0.5\text{--}1^\circ$. No minimum fixation duration criteria were used (although the temporal resolution does introduce an uncertainty of 20 ms to our determination of fixation durations). We did not include any account of blinks in our analyses.

Throughout the analyses of this experiment, performance is reported as a difference score, where chance is subtracted from raw performance scores. Chance was calculated for each question by using the

frequency of correct responses to that question in the control experiment. Thus a difference score of zero would indicate that participants performed at chance in the question, whereas a positive score would indicate above-chance performance. This approach allowed us to assess performance in the main experiment relative to schema-based guesses in the control experiment.

Eye movement data and performance were used to characterize the two aspects of object information retention discussed in the Introduction: its uptake and its stability. Uptake was characterized by relating performance in each question to the number of times the tested object was fixated during the trial. One possible limitation of using the number of fixations on an object as the dependent variable is that it may not be valid to treat fixations of different durations as equivalent units for extracting information. We therefore repeated our analyses using total fixation time (the sum of the durations of all fixations on the target object) on the object as the dependent variable. Stability was described by comparing performance to the number of fixations that occurred between the final fixation on the tested object and the end of the trial.

Performance on each of the five question categories (presence, shape, colour, position, and relative distance) was analysed independently using a one-way repeated measures analysis of variance (ANOVA) with polynomial trend analysis. The effect size measure of partial η^2 is reported for main effects and linear trends. Missing values were replaced by series means. Outlying data points were removed if they were more than two standard deviations from the series mean. Trend analysis indicates whether any trends are present, but does not indicate the slope of the trend or the effect size, which could be derived using analyses such as multiple regression. However, comparing slopes or effect sizes would not be valid. For this reason we feel that trend analysis, which indicates the presence or absence of a trend, is sufficient and appropriate for our purposes. To test whether information was encoded during viewing, *a priori* planned comparisons were conducted using one-sample *t* tests to compare performance to chance levels for objects fixated 0, 1, 2, 3, or 4+ times. These *t* tests were not conducted for the analysis of information stability because stability is characterized by whether performance changed with increasing number of intervening fixations, not the performance scores per se.

Results

We found no practice effect over the course of the experiment. One-way repeated measures ANOVA with linear trend analysis showed that while there was a main effect of trial number on performance, $F(5, 22) = 5.74$, $p = .002$, partial $\eta^2 = .566$, there was no significant linear trend in performance across the six trials of the experiment, $F(1, 26) < 0.01$, $p > .999$, partial $\eta^2 < .001$. Hence participants did not show any overall improvement or decline in performance over the course of the experiment.

It was possible that object information encoded during viewing might have faded over the course of filling in the questionnaire. We tested this by comparing performances in the first, second, and third sections of the questionnaires (see Method). One-way repeated measures ANOVA with linear trend analysis showed no significant main effect, $F(2, 25) = 2.45$, $p = .107$, partial $\eta^2 = .164$, or linear trend in performance across the three sections of the questionnaires, $F(1, 26) = 1.82$, $p = .189$, partial $\eta^2 = .065$; hence we found no evidence that performance decreased over the course of completing the questionnaire.

Inspection behaviour

Fixations had a median duration of 220 ms and a mean of 284 ms ($SD = 190$ ms, $N = 2,259$). This corresponded to an average of 14.49 fixations per room ($SD = 4.12$). Saccades had a median amplitude of 8° and a mean of 11.2° ($SD = 9.74$, $N = 2,260$).

The frequency with which target objects were fixated in the experiment can be assessed by considering whether each participant fixated a particular target object during each trial. We can then quantify the number of participants that fixated each target object in each room and use this to produce an average number of participants that viewed the target objects. Overall, target objects were fixated at least once by an average of 13.4 participants. All target objects were fixated by at least two participants.

Information uptake

We can examine any retention of information by looking at the effect of the number of times the target item was fixated during viewing on performance. This provides a measure of the extent of information accumulation across fixations. Fixations on target objects did not need to be consecutive to be included in the analyses below. Information uptake and retention are measured in two ways. We can determine whether information is encoded by comparing performance difference scores in each cell (0, 1, 2, 3, 4+ fixations of an object) to chance (with zero representing no difference from chance and positive values indicating above-chance performance), using the results of the planned one-sample t tests. We can characterize any information accumulation from multiple fixations by using the results of the one-way ANOVAs with linear trend analysis on the data for objects fixated 1, 2, 3, and 4+ times. It is not appropriate to include nonfixated objects in this analysis because we are interested in the accumulation of information across fixations. Higher order trend analyses were also conducted, but the results of these are only reported if significant.

There are clear qualitative differences between the five types of object property tested (Figure 3). For object presence, data are only available for target objects that were not fixated, fixated once, or fixated twice during viewing. Data are not included or analysed for the 3 and 4+ fixation cells because there were not enough observations for these cells (only 2 for objects fixated 3 times and 1 for objects fixated 4+ times). This lack of data is because target objects in the presence questions happened not to have been fixated more than twice on most occasions. A single-factor ANOVA was carried out with two levels of fixation (1 and 2 fixations on the object). There was no significant main effect of the number of fixations on the target objects, $F(1, 25) = 2.35$, $p = .138$, partial $\eta^2 = .086$. Whether information was encoded during viewing was examined by using one-sample t tests to compare observed performance to chance. Performance was significantly above chance for objects that were not fixated, $t(25) = 3.10$, $p = .005$, fixated once, $t(25) = 5.26$, $p < .001$, or fixated twice, $t(25) = 3.47$, $p = .002$.

For questions about object shape, data were excluded from the analysis for objects fixated four or more times as there were only four observations for this cell, so the ANOVA had three levels: 1, 2, and 3 fixations. One-way ANOVA with linear trend analysis showed neither a significant main effect of the number of fixations, $F(2, 24) = 1.32$, $p = .287$, partial $\eta^2 = .099$, nor a significant linear trend, $F(1, 25) = 0.50$, $p = .485$, partial $\eta^2 = .020$. Performance was not significantly above chance for objects that were not fixated or for objects that were fixated one, two, or three times ($t < 1.9$, ns for all comparisons; Bonferroni corrected).

Data were excluded from the object colour analyses for objects fixated four or more times as there were only four observations for this cell; thus the ANOVA had three levels (1, 2, and 3 fixations of the target object). There was no significant main effect, $F(2, 24) = 1.83$,

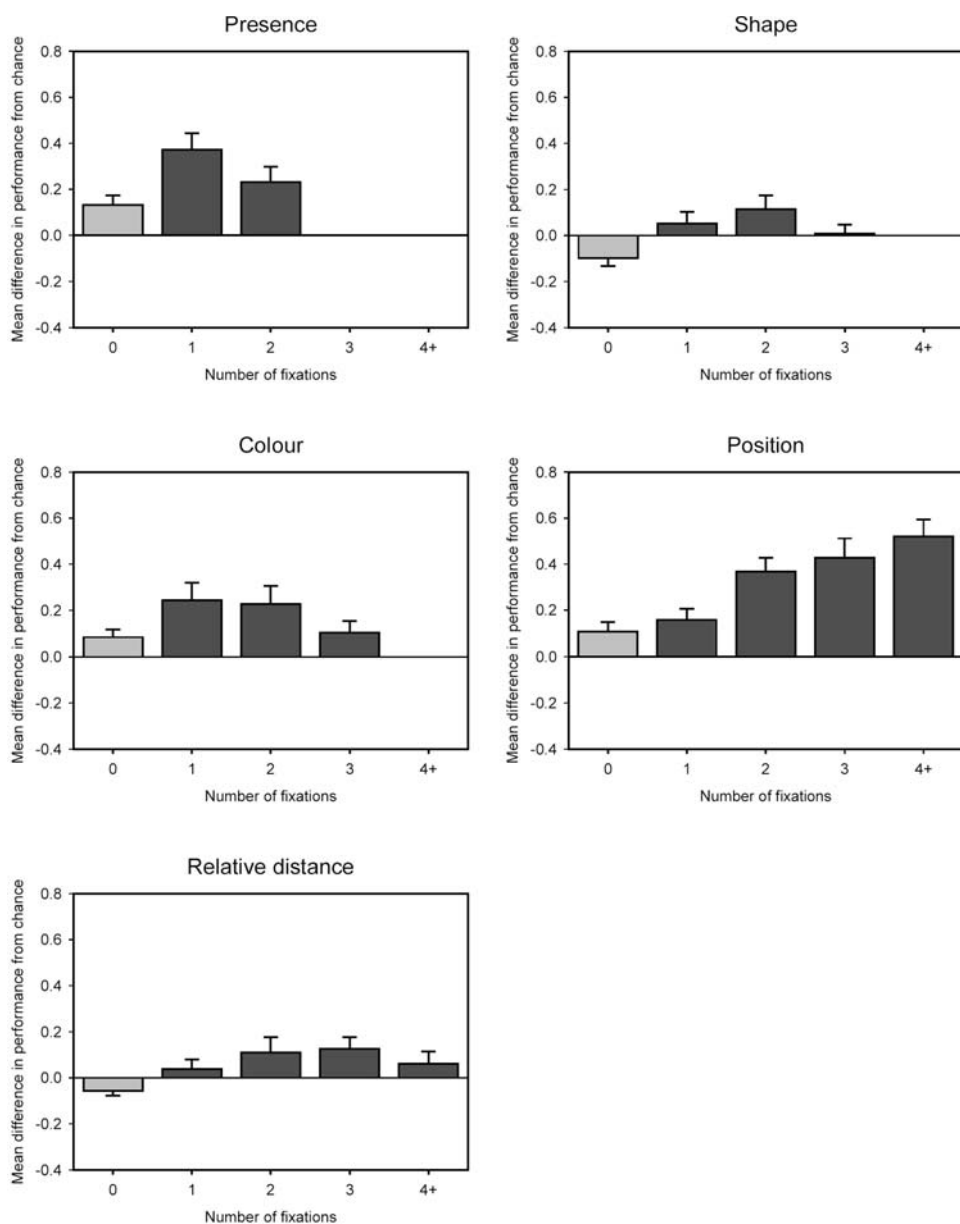


Figure 3. Performance difference scores ($\pm 1 SE$) for each of the five question types tested in Experiment 1 as a function of the number of fixations on the target objects. A difference score of zero represents no difference from chance, and positive scores indicate above-chance performance. The zero fixations case is shaded lighter because it was not included in the ANOVA analyses of information accumulation.

$p = .182$, partial $\eta^2 = .133$, and no significant linear trend, $F(1, 25) = 2.93$, $p = .099$, partial $\eta^2 = .105$. Performance was significantly above chance for objects that were fixated once, $t(25) = 3.23$, $p = .003$, or twice, $t(25) = 2.92$, $p = .007$, during viewing ($t < 2.4$, ns if not fixated or fixated 3 times; Bonferroni corrected).

For questions testing object positions, there was a significant main effect of the number of fixations, $F(3, 23) = 9.51$, $p < .001$, partial $\eta^2 = .554$, with a significant linear trend of increasing performance with increasing number of fixations on the object, $F(1, 25) = 20.29$, $p < .001$, partial $\eta^2 = .448$. Performance was significantly above chance for objects fixated once, $t(25) = 3.45$, $p = .002$, twice, $t(25) = 6.23$, $p < .001$, three times, $t(25) = 4.98$, $p < .001$, or four or more times, $t(25) = 6.91$, $p < .001$.

For questions testing relative distances between objects there was no significant main effect, $F(3, 23) = 1.07$, $p = .381$, partial $\eta^2 = .123$, and no significant linear trend, $F(1, 25) = 0.10$, $p = .752$, partial $\eta^2 = .004$. Performance was not significantly above chance for objects that were not fixated or for objects that were fixated any number of times ($t < 1.6$, ns for all comparisons; Bonferroni corrected).

In order to account for the possible limitation of treating fixations of different durations as equivalent in the above analyses (see above), data for information uptake were reanalysed using total fixation time on the object as the dependent variable. We again used one-way ANOVAs with trend analysis for objects fixated for 0, 1–400, 401–800, 801–1,200, and 1,201+ ms. The results of this reanalysis using total fixation time were largely the same as those reported above using the number of fixations on objects; this is perhaps unsurprising given that fixation number and total fixation time are likely to be highly correlated. As before, no main effects or trends were found for questions testing presence, shape, and relative distance information. The same pattern as before was found for object position information; there was a main effect of total fixation time, $F(3, 23) = 19.17$, $p < .001$, partial $\eta^2 = .615$, and a significant linear trend of increasing performance with time spent fixating the object, $F(1, 25) = 30.66$, $p < .001$, partial $\eta^2 = .551$. The only question type that showed a different result in this reanalysis was colour; here a significant main effect of fixation time was found for questions testing colour, $F(2, 24) = 8.99$, $p = .001$, partial $\eta^2 = .428$, with a significant linear trend of decreasing performance with increasing total time spent fixating an object, $F(1, 25) = 11.11$, $p = .003$, partial $\eta^2 = .308$.

Information stability

In considering the stability of object descriptions, it is only appropriate to look at those forms of information that showed evidence for encoding in the above analyses. We found evidence for encoding and retention of information about object presence, colour, and position, but there was little evidence for retention of shape or relative distance information. We can study the stability with which information is retained by comparing performance to the number of intervening fixations between the final fixation on the object and the end of the trial (after Hollingworth & Henderson, 2002). Data were collapsed for trials in which there were 1 or 2 intervening fixations and also for trials in which there were 3 or 4 intervening fixations. This collapsing was performed because there were some missing values in the individual cells; using this method therefore reduced the noise in the data. We assessed stability by using one-way ANOVAs with linear trend analysis on the data for objects with 0, 1–2, 3–4,

and 5+ intervening fixations. No *t* tests were conducted for the analysis of information stability because stability is characterized by whether performance changed with increasing number of intervening fixations, not the performance scores per se.

Figure 4 plots performance (in terms of difference from chance) as a function of the number of intervening fixations. There was a significant main effect of intervening fixations for object presence questions, $F(3, 23) = 93.28$, $p < .001$, partial $\eta^2 = .924$, with a significant linear trend of decreasing performance as the number of intervening fixations increased, $F(1, 25) = 22.92$, $p < .001$, partial $\eta^2 = .478$. A significant main effect was found for questions testing object colour, $F(3, 23) = 3.71$, $p = .026$, partial $\eta^2 = .326$, with a significant linear trend of decreasing performance as the number of intervening fixations increased, $F(1, 25) = 6.56$, $p = .017$, partial $\eta^2 = .208$. There was no significant main effect for object position questions, $F(3, 23) = 1.56$, $p = .482$, partial $\eta^2 = .020$, and no significant linear trend, $F(1, 25) = 0.51$, $p = .371$, partial $\eta^2 = .032$.

Discussion

The results from Experiment 1 demonstrate that some object properties can be encoded and retained while we view real-world scenes, when we are required to attend to multiple object properties concurrently. Information about object presence in the scene, the colour of objects, and positions of objects all show evidence that they are encoded and retained. However, not all object properties are retained: We found no evidence to suggest that object shape information or the relative distances between objects were encoded and retained.

For the three types of information that were encoded and retained, two different patterns were found. For object position information, performance clearly increased with increasing number of fixations, or total fixation time, on an object. This result shows that information about object positions in real-world scenes accumulates from multiple fixations. Object presence and colour information did not show this same pattern of increasing performance. Increasing numbers of fixations, or total fixation time, did not appear to increase performance for either of these object properties. This result implies that these types of information are not accumulated during revisits to an object. While our data imply that presence and colour information are not accumulated, these findings require further exploration because of the limited data available in some cells of the analyses (see Results).

Looking at performance on trials when the target object was not fixated can provide an indication of whether fixation of an object is required to gather information about it. For object colour and presence information, performance was above chance in trials when the target objects were not fixated. This result suggests that these two forms of information can be derived without direct fixation of an object. Conversely, performance for object position questions was not significantly different from chance when the target objects were not fixated. Hence, it appears that direct fixation of an object is required to extract meaningful position information.

The three types of information that we found to be encoded did not show the same patterns of stability. Object presence showed a clear recency effect, with performance decreasing markedly as the number of intervening fixations between the last fixation of the object and the end of the trial increased. Object colour showed a significant recency effect, and hence it appears that colour information is retained transiently in a similar manner to object presence.

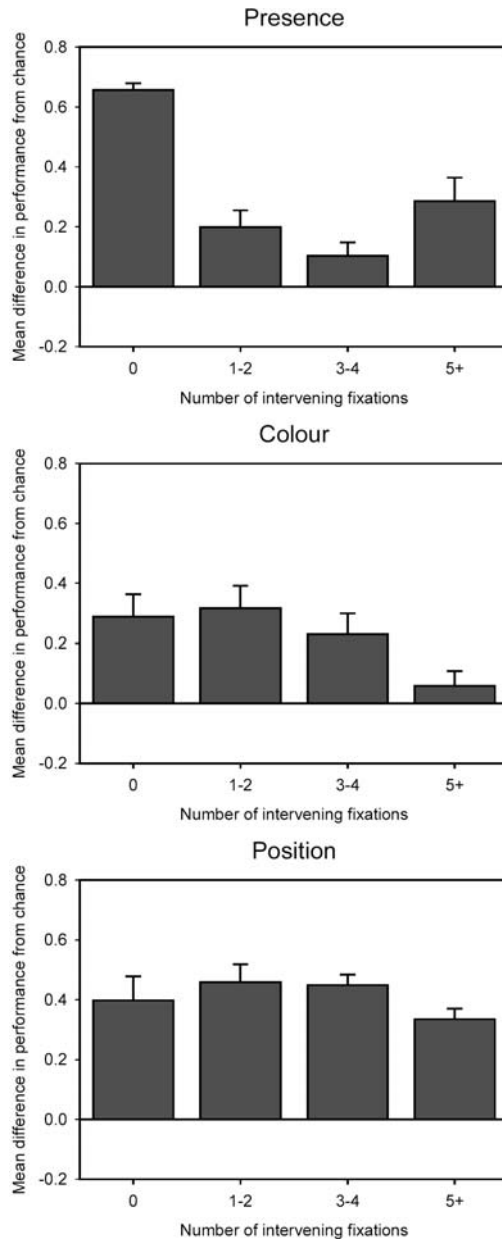


Figure 4. Performance difference scores ($+1\ SE$) for the three types of information encoded in Experiment 1 (see Figure 3 and text) as a function of the number of intervening fixations between the last fixation of the target object and the end of the trial.

For object position information, there was no observed change in performance with intervening fixations. This result shows that object position information is not transient, but is retained stably once encoded. Our data for information stability do not clearly argue for an overall stable or transient retention of encoded object properties; rather they argue that stability differs for different forms of encoded information.

In interpreting our data and comparing these to existing studies of object information encoding and retention, we must distinguish two possible factors and consider their relative contribution to the observations. First, this experiment differs from previous research by testing several types of information concurrently, whereas the majority of previous studies consider object properties one (or sometimes two) at a time. Second, this experiment was conducted in a real-world setting, whereas previous studies typically use computer viewing paradigms. Both of these factors may have contributed to the observed patterns of information encoding and stability. To differentiate between these two possibilities, in Experiment 2 we repeat Experiment 1, but with photographic images of the rooms displayed on a computer monitor.

EXPERIMENT 2

In this experiment participants viewed computer-displayed photographic images of the real-world scenes used in Experiment 1. This experiment allows us to compare and contrast real-world and computer-viewing experimental approaches to the study of object and scene memory. In addition Experiment 2 may strengthen the evidence for differential encoding, accumulation, and retention of information from natural scenes under conditions of concurrent processing.

Method

Participants

A total of 32 participants (mean age 22.1 years, $SD = 6.2$) took part in this experiment. None of these participants took part in Experiment 1 or the control experiment. All had normal or corrected-to-normal vision.

Procedure

Participants viewed photographic images of the six rooms used in Experiment 1. The content and layout were exactly as had been viewed by participants in the first experiment. Images were displayed on a 21" colour monitor positioned at a viewing distance of 60 cm. The images subtended approximately $40^\circ \times 30^\circ$ of the participant's visual field. Each trial started with a central fixation target and was followed by the presentation of the image for 5 s. After 5 s the image disappeared, and the participants were presented with questionnaires identical to those completed by participants in Experiment 1.

Eye movement recording

Eye movements were monitored during image viewing using the SR Research Ltd. EyeLink II eye tracker, which samples eye position data at 500 Hz. Eye position data were collected binocularly and analysed for the eye that produced the better spatial accuracy. A 9-point target display was used for

calibration of eye position. A second 9-point display was used to validate the calibration and return the mean spatial accuracy of the eye tracker calibration. Further 9-point validations of the calibration were carried out at the start of each experimental trial. If the validation showed that the spatial accuracy of the eye tracker had deteriorated to worse than $\pm 0.5^\circ$, the eye tracker was recalibrated as described above. In this study, the mean spatial accuracy of the eye tracker calibration was 0.31° ($SD = 0.08^\circ$).

Analysis

Analysis was carried out off-line after completion of the experiment. Extraction of gaze position was carried out using software supplied with the EyeLink II eye tracking system. Saccade detection required a deflection of greater than 0.1° , with a minimum velocity of $35^\circ/s$ and a minimum acceleration of $9,500^\circ/s^2$, maintained for at least 4 ms. No minimum fixation duration criteria were imposed. Fixations that were ongoing when the trial terminated were included as valid fixations. We did not include any account of blinks in our analyses. Gaze position was compared to the position of objects tested in the questionnaires. As in Experiment 1, an object was judged to have been fixated if the centre of gaze fell within 1.5° of it. Six responses (from 5 participants) were excluded because the participants failed to answer these questions. These exclusions left 1,722 responses from 192 trials for analysis.

Following the method used in Experiment 1, performance difference scores, where chance is subtracted from raw performance scores, were calculated.

Eye movement data and performance were used to investigate the same two issues as those in the first experiment—information uptake and information stability.

Results

We found no practice effect over the course of the experiment. One-way repeated measures ANOVA with linear trend analysis showed that there was a significant main effect of trial number, $F(5, 27) = 4.03$, $p = .007$, partial $\eta^2 = .428$, but no significant linear trend in performance across the six trials of the experiment, $F(1, 31) = 0.05$, $p = .823$, partial $\eta^2 = .002$. Participants showed no overall improvement or decline in performance over the course of the experiment.

One-way repeated measures ANOVA with linear trend analysis showed no significant main effect of questionnaire section, $F(2, 30) = 1.26$, $p = .297$, partial $\eta^2 = .078$, and no trend in performance across the three sections of the questionnaires, $F(1, 31) = 1.31$, $p = .261$, partial $\eta^2 = .041$; thus we found no evidence that performance decreased over the course of completing the questionnaire.

Inspection behaviour

Fixations had a median duration of 212 ms and a mean of 233 ms ($SD = 119$, $N = 3,345$). This corresponded to an average of 17.53 fixations per room ($SD = 2.62$). Saccades had a median amplitude of 4.77° and a mean of 6.62° ($SD = 6.05$, $N = 3,174$).

The frequency with which target objects were fixated in the experiment can be assessed by considering whether each participant fixated a particular target object during each trial. We can then quantify the number of participants that fixated each target object in each room and use this to produce an average number of participants that viewed the target objects.

Overall, target objects were fixated at least once by an average of 22.89 participants. All target objects were fixated by at least four participants.

Information uptake

Figure 5 plots the performance by participants (as difference from chance level) as a function of the number of fixations on an object. There are clear qualitative differences between the patterns for the five question types. For presence questions, there was a significant main effect of the number of fixations on an object, $F(3, 29) = 13.04$, $p < .001$, partial $\eta^2 = .574$. There was no significant linear trend, $F(1, 31) = 0.66$, $p = .422$, partial $\eta^2 = .021$, but there was a significant quadratic trend of changing performance with increasing number of fixations on an object, $F(1, 31) = 26.44$, $p < .001$, partial $\eta^2 = .460$. One-sample t tests showed that performance was above chance for objects that were not fixated, $t(31) = 2.94$, $p = .006$, fixated once, $t(31) = 3.82$, $p = .001$, or fixated four or more times, $t(31) = 3.99$, $p < .001$, but not for objects fixated two or three times ($t < 1.8$, ns , for both; Bonferroni corrected).

For shape information, there was a significant main effect, $F(3, 29) = 5.21$, $p = .005$, partial $\eta^2 = .350$. There was no significant linear trend, $F(1, 31) = 3.15$, $p = .086$, partial $\eta^2 = .092$, but there was a significant quadratic trend, $F(1, 31) = 10.54$, $p = .003$, partial $\eta^2 = .254$. Performance was not significantly above chance for objects that were not fixated, or fixated any number of times ($t < 0.6$ for all, ns ; Bonferroni corrected).

Performance for questions testing colour showed no significant main effect, $F(3, 29) = 0.94$, $p = .433$, partial $\eta^2 = .089$, and no linear trend, $F(1, 31) = 1.86$, $p = .183$, partial $\eta^2 = .056$. Performance was significantly above chance for objects that were not fixated, $t(31) = 3.14$, $p = .004$, fixated twice, $t(31) = 5.02$, $p < .001$, or fixated four or more times, $t(31) = 2.74$, $p = .010$. Performance was no different from chance for objects that were not fixated or fixated twice ($t < 2.4$, ns ; Bonferroni corrected).

For questions testing object position information, there was a main effect of the number of fixations, $F(3, 29) = 7.11$, $p = .001$, partial $\eta^2 = .424$, and a significant linear trend of increasing performance with increasing number of fixations on the object, $F(1, 31) = 7.69$, $p = .009$, partial $\eta^2 = .199$. Performance was significantly above chance for objects fixated two times, $t(31) = 4.51$, $p < .001$, three times, $t(31) = 3.96$, $p < .001$, or four or more times, $t(31) = 4.20$, $p < .001$. Performance was no different from chance for objects that were not fixated or fixated only once ($t < 2.5$, ns ; Bonferroni corrected).

Relative distance questions showed no significant main effect, $F(3, 29) = 2.59$, $p = .072$, partial $\eta^2 = .211$, and no significant linear trend, $F(1, 31) = 0.43$, $p = .515$, partial $\eta^2 = .014$. Performance was not significantly above chance for objects that were not fixated or those that were fixated any number of times ($t < 1.2$ for all, ns ; Bonferroni corrected).

Reanalysis of the data, using total time spent fixating an object as the dependent variable, again showed largely the same patterns as those shown by the above analyses. No linear or higher order trends were found for questions testing object presence, shape, colour, or relative distances, and no main effect of total fixation time was found for questions testing object colour. In contrast to the analyses using fixation number, no main effects of total fixation time upon performance were found for questions testing presence or shape information. There was a significant main effect of time spent fixating an object for questions testing the relative

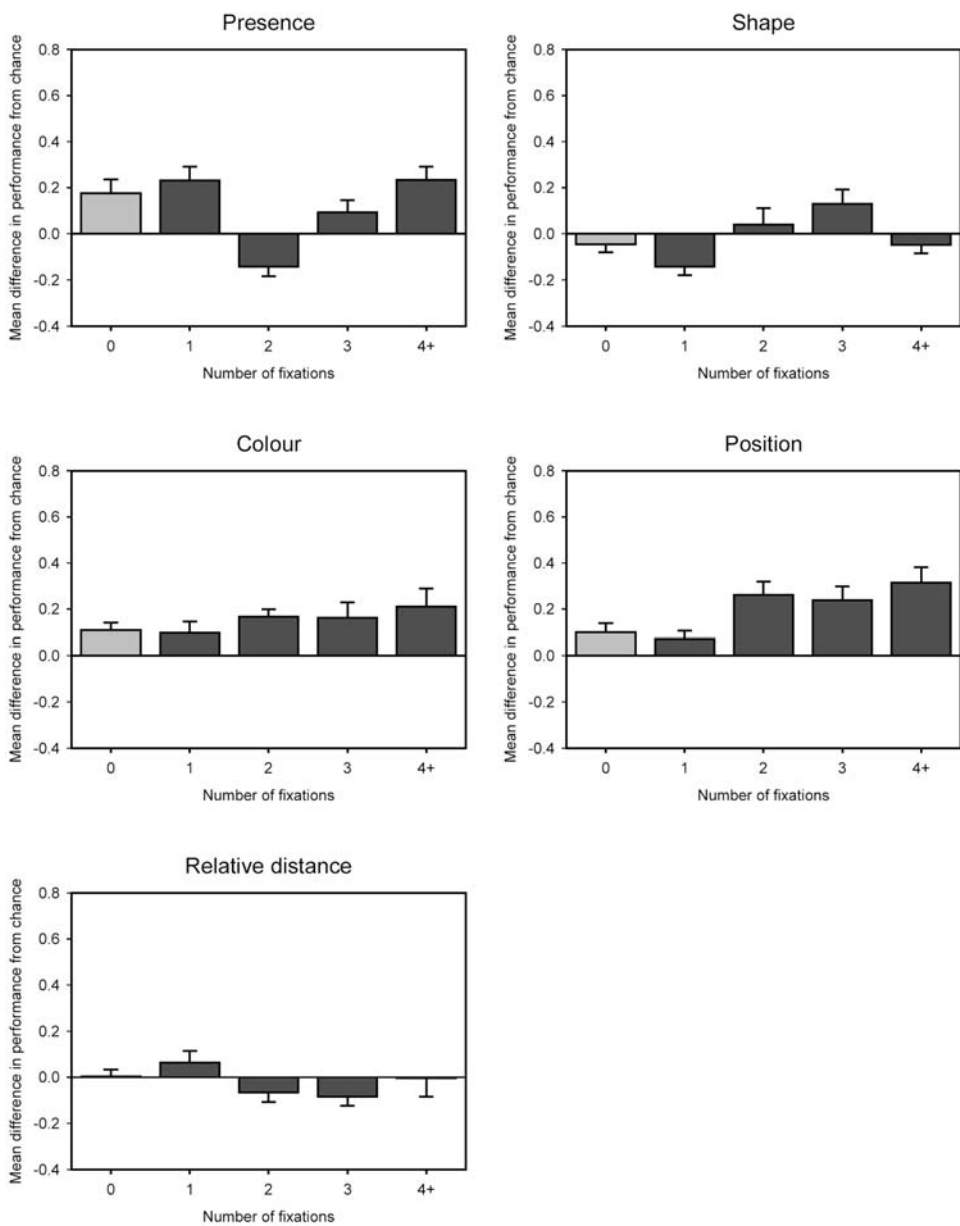


Figure 5. Performance difference scores ($+1 SE$) for each of the five question types tested in Experiment 2 as a function of the number of fixations on the target objects. A difference score of zero represents no difference from chance, and positive scores indicate above-chance performance. The zero fixations case is shaded lighter because it was not included in the ANOVA analyses of information accumulation.

distances between objects, $F(3, 29) = 7.13$, $p = .001$, partial $\eta^2 = .425$. As in the above analyses, there was a significant main effect of total fixation time upon performance for questions testing object position, $F(3, 29) = 10.54$, $p < .001$, partial $\eta^2 = .522$, with a significant linear trend of increasing performance as the total time spent fixating an object increased, $F(1, 31) = 7.96$, $p = .008$, partial $\eta^2 = .204$.

Information stability

It is only legitimate to look for any patterns in stability for the three object properties that demonstrated retention of information: presence, colour, and position.

Figure 6 plots performance as a function of the number of intervening fixations between the last one on the target object and the end of the trial. There was a main effect of the number of intervening fixations for questions testing object presence, $F(3, 29) = 6.49$, $p = .002$, partial $\eta^2 = .402$, and a significant linear trend of decreasing performance with increasing number intervening fixations, $F(1, 31) = 8.47$, $p = .007$, partial $\eta^2 = .215$. For colour information, there was a significant main effect, $F(3, 29) = 3.75$, $p = .022$, partial $\eta^2 = .279$, but no significant linear trend, $F(1, 31) = 0.94$, $p = .761$, partial $\eta^2 = .003$. There was a significant main effect of the number of intervening fixations for position information, $F(3, 29) = 6.22$, $p = .002$, partial $\eta^2 = .391$, and a significant linear trend of decreasing performance with increasing number of fixations between the last fixation of the target object and the end of the trial, $F(1, 31) = 14.91$, $p = .001$, partial $\eta^2 = .325$.

Discussion

The results of Experiment 2, when observers viewed computer-displayed images of real scenes, are largely supportive of those found in Experiment 1, for viewing real scenes. We therefore confirm that under conditions of concurrent information processing, while viewing natural scenes, object presence, colour, and position information are extracted and retained. Conversely, there was no evidence to suggest that object shapes and the relative distances between objects were retained.

The same patterns of information encoding were observed for object presence, colour, and position information as were found in Experiment 1. Both object presence and colour information showed above-chance performance in trials when the target objects were not fixated, whereas object position was not different from chance when the target objects were not fixated. Hence object presence and colour information can be encoded without direct fixation, whereas object position cannot. Position information accumulates from multiple fixations, with performance increasing with increasing number of fixations, or total fixation time, on the target object. Conversely, there was no evidence for accumulation of presence or colour information; both of these types of object description were unaffected by the number of times the target item was fixated or for how long it was fixated.

While there were no significant linear trends in performance with increasing number of fixations on the target objects for presence or shape information, there were quadratic trends for both of these. The quadratic trend found for shape information is not easy to explain given that performance was not significantly different from chance for objects fixated any number of times. The importance of these trends is also questioned by the fact that no such

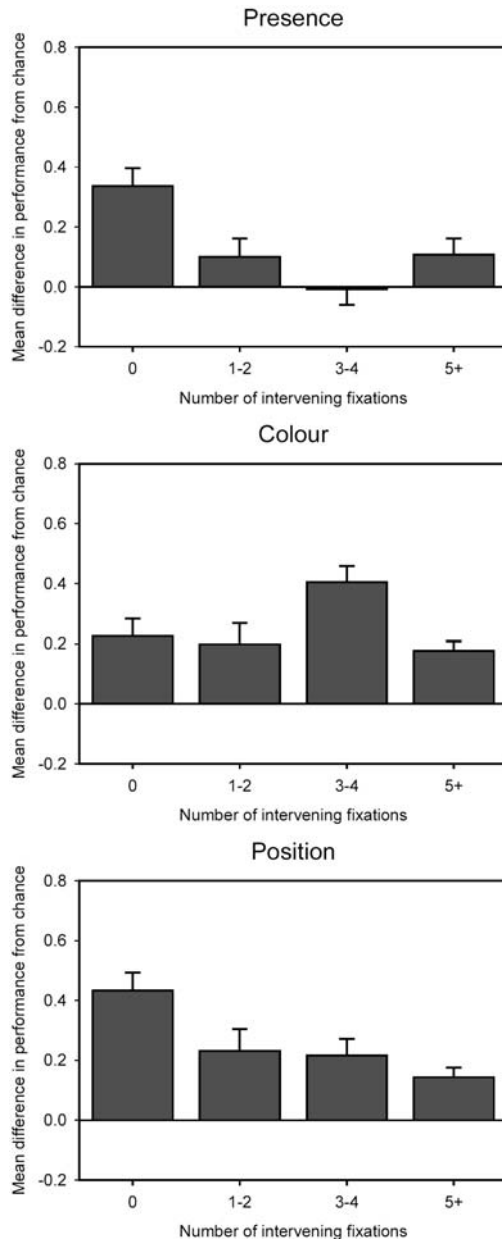


Figure 6. Performance difference scores ($\pm 1 SE$) for the three types of information encoded in Experiment 2 (see Figure 5 and text) as a function of the number of intervening fixations between the last fixation of the target object and the end of the trial.

trends were found in Experiment 1, nor were quadratic trends found for these question types when the data were reanalysed with respect to the time spent fixating the object rather than the number of fixations on the object. Consequently, while the quadratic trends may reflect subtle patterns of extraction for presence and shape information, we cannot be certain of their implications given the present data.

The stability of encoded presence information in this experiment showed the same transience as that found in Experiment 1: Performance decreased as the number of intervening fixations between the last fixation of the target object and the end of the trial increased. For the other two types of encoded information, however, the patterns of stability found in Experiment 2 were different from those found in Experiment 1. Object colour information showed no significant influence of the number of intervening fixations between the last fixation of the target object and the end of the trial, suggesting that colour was retained stably once encoded. Position information showed a significant decline in performance with increasing number of intervening fixations, suggesting that position information is retained transiently. These two patterns are the opposite of those observed in Experiment 1.

Our data for information encoding and accumulation in Experiment 2 are again inconsistent with our initial hypotheses; we expected to find retention and accumulation of all types of information and to find the same patterns of stability (or transience) for information once retained. However, they are consistent with the results of Experiment 1. In terms of information stability, the results of Experiment 2 are broadly in line with those of Experiment 1, in that they imply differential retention of object properties, and both demonstrated that object presence information was transient. However, the observed patterns of stability for colour and position information were different from those found in Experiment 1.

GENERAL DISCUSSION

The two experiments reported in this study were designed to investigate the encoding and subsequent stability of object descriptions from real-world and computer-displayed scenes. The experiments allowed us to determine which object properties are encoded and to characterize the manner of this encoding from fixations. A crucial component of the experimental approaches was testing several types of object information concurrently. This was done in order to more closely approximate likely processing demands under everyday conditions, when we are likely to routinely attend to multiple object properties at the same time. In this way we hoped that our observed patterns of encoding might be more ecologically valid than approaches where single types of object property are tested. An important comparison, and further test of the ecological validity of the findings, was between encoding and retention when viewing real three-dimensional scenes (Experiment 1) and computer-displayed photographic images of the same scenes (Experiment 2).

Three aspects of the experimental findings are now considered. First is the way in which the data can be used to characterize the two aspects of the formation of object descriptions that were considered in the Introduction: the uptake of information and the stability with which it is subsequently retained. Second, findings from the two experiments are compared in order to consider the similarities and differences between information extraction from real-world scenes and computer-displayed images of the same scenes. Finally, the experimental findings are related to existing views of object memory encoding and retention.

Information uptake

We found consistent patterns of encoding in both experiments. Object presence, colour, and position information were encoded and retained. Encoding and retention of object properties are consistent with a large body of recent evidence (e.g., Henderson, 1994; Henderson & Hollingworth, 1999; Henderson & Siefert, 1999; Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002; Melcher, 2001; Tatler et al., 2003). However, not all object properties were retained: There was no evidence for encoding of object shape and the relative distances between objects. Retention of some object properties but not others is consistent with Henderson's (1994) suggestion that *object files* might preserve abstract information such as identity, but not detailed information such as form. After our consideration of a range of object properties, therefore, the data imply that retained object descriptions do not encompass all object properties, but rather are selective, comprising particular object properties but not others.

For those types of information that were encoded, the manner of encoding from fixations varied. Position information was accumulated from multiple fixations on the target objects, with performance increasing with increasing number of fixations, or total fixation time, on the target. Accumulation of position information is consistent with Hollingworth and Henderson's (2002) finding that object properties accumulate with fixation time during viewing (although they did not test position information in their experiments). Tatler et al. (2003) also found that object properties accumulate over the course of several seconds of viewing natural scenes. However, object presence and colour information did not accumulate in either of our experiments. Irwin and Andrews (1996) also found that identity information does not accumulate across fixations using a partial report paradigm for letter array stimuli. They interpreted this to suggest that object properties are generally not accumulated across fixations. Our observed lack of accumulation of object presence information is in contrast to Hollingworth and Henderson's (2002) general suggestion that object properties (including object token information) accumulate but, interestingly, is consistent with the findings of their Experiment 3, where no evidence for accumulation was found for two-alternative forced-choice questions testing object token information.

Our experiments demonstrate that patterns of encoding of individual object properties ought not to be generalized to the encoding of all object properties. Differential encoding of object properties has been suggested (e.g., Carlson-Radvansky & Irwin, 1995; Irwin, 1992a). Irwin suggested that object identity can be encoded without position information to index the object file to the scene (Irwin, 1992a). Carlson-Radvansky and Irwin (1995) suggested that object structural descriptions can be independent of position information.

Information stability

In both experiments we found that object presence information was transient, with performance decreasing with increasing number of intervening fixations between the last fixation of the target object and the end of the trial. The general pattern of transience was for performance to diminish over the course of several fixations, rather than to drop to chance after the end of the last fixation on the target object. Thus our data do not support the suggestion that information is limited to the current target of attention (as suggested by, e.g., O'Regan, 1992;

O'Regan & Noë, 2001; Rensink, 2000a, 2000b; Rensink et al., 1997). Rather our observed patterns of transience were consistent with those suggested by Irwin's object file theory (Irwin, 1992a, 1992b; Irwin & Andrews, 1996), which suggests that information would be expected to decay over the course of several fixations following the last fixation of a target object. Irwin and Zelinsky (2002) found this pattern of transience for object token information (which is a form of identity information), and their time course of decay was similar to that which we found in our two experiments, with performance dropping markedly over the next two to three fixations after leaving the tested object.

While our two experiments were consistent in terms of the observed transience of object presence information, for object colour and position information, opposite patterns of stability were found in the two experimental settings. In Experiment 1, we found that object colour information was transient, but position information was retained stably. Conversely in Experiment 2, object colour information was retained stably, but position was transient. The large effect sizes for colour in Experiment 1 and position in Experiment 2 suggest that these findings of transience cannot be discounted. Neither can the data for position in Experiment 1 or for colour in Experiment 2 be seen to approach significance or show any trend of transience. It therefore appears that for colour and position information different patterns of information retention occur when viewing real-world scenes than when viewing computer-displayed images of those scenes. Hollingworth and Henderson (2002) found that their tested object properties were retained stably, with no evidence of transience. Thus there is support for our suggestion that under some conditions object properties can be retained stably.

While our data for the stability with which information is retained once encoded are somewhat less consistent between the two experiments than our data for information uptake, they again clearly argue for some degree of independence of the object properties. As before, findings from single-object properties cannot be generalized to all object properties.

Ecological validity of computer-viewing approaches

There are a number of ways in which computer images of scenes differ from the real world. Overall these differences result in a richer perceptual input in the real world when compared to computer presentation of the same scenes. For example, the real-world scene has an extended field of view, real depth, and a greater range of brightnesses. These differences in the perceptual input may have real consequences for the encoding and stability of the information about the scene. Comparing our observations from the two experiments allows us to contrast viewing strategies and information retention under these two viewing conditions.

In terms of inspection behaviour, we found very similar median fixation durations in the two settings. However, for real scenes, the mean fixation duration was higher, and the average number of fixations on each scene was fewer than for computer-displayed scenes, despite equal viewing times. This suggests that there may have been more of the longer fixations when viewing real scenes than when viewing two-dimensional scenes. Saccades tended to be of greater amplitude when viewing real scenes than when viewing computer-displayed scenes. This may reflect the different scales of the two scenes; computer-displayed scenes subtended approximately 40° by 30° , whereas the real scenes were much larger.

The high degree of consistency between our findings in the two experiments for information uptake strongly suggest that the strategies and mechanisms for encoding information

from real-world scenes in Experiment 1 and computer-displayed scenes in Experiment 2 were very similar. This is a strong argument that experimental studies of object and scene perception in general using computer-displayed scenes (as is often the case in the literature) can provide valid insights into the operation of the perceptual system under real-world conditions.

The data for the stability with which information is retained once encoded are less consistent between the two experiments. Both found that presence information was transient, but were inconsistent in terms of the stability of colour and position information. This discrepancy might imply that while encoding strategies are the same in the two experimental settings, the object description that is constructed from the encoded properties differs in the real world from the computer-viewing setting. If this is the case then differential information stability in real-world settings and computer-viewing experiments should be considered when using computer-viewing paradigms to characterize aspects of object and scene perception.

The consistency between Experiments 1 and 2 appears to suggest that computer-based viewing paradigms may be sufficient for investigating how the visual system encodes and retains information in the real world. However, looking at the effect sizes reveals much stronger effects in the real-world setting than in the computer-viewing experiment (see particularly the linear trends for accumulating position information and for the transience of presence information). Therefore subtle effects that exist in the real world might be missed if research in this area relies solely on computer-based paradigms. While the two experimental settings were roughly comparable in the present study, it should be remembered that the task was a simple viewing and memory task. There was no locomotion of the participants within the environment or manipulation by the participants of the content of the scenes; both of these typically occur in real-world situations. Such dynamic interaction between the observer and the scene are not often possible for computer-based viewing paradigms. Therefore, to investigate the dynamic aspects of vision and the encoding and retention of object information, real-world approaches are essential.

Existing models of information retention

There have been numerous suggested characterizations of the encoding and retention of object information. We now consider how our data relate to some of the most prominent hypotheses of object information encoding and retention.

There exist three particularly prominent accounts of transsaccadic information encoding and retention. First, Rensink's (2000a, 2000b) *coherence theory* proposes that object descriptions comprise attended object properties for the current focus of attention. Each object description dissolves as soon as attention is withdrawn from the object. Two aspects of our data clearly argue against this proposition. First, we find evidence for accumulation of object position information from multiple (not necessarily consecutive) fixations of objects. Second, we find that object properties either decay over a number of fixations or are stable, whereas coherence theory predicts that object properties should be lost as soon as the last fixation of a target object is terminated.

Irwin and colleagues (Carlson-Radvansky & Irwin, 1995; Irwin, 1992a, 1992b; Irwin & Andrews, 1996) have argued that information is extracted from fixations and held in visual short-term memory as object files (see Kahneman & Treisman, 1984). Around 3–4 of these

object files can be retained simultaneously, and they comprise descriptions of the properties of recently attended objects. Object files endure after attention is withdrawn from an object but are replaced by new object files when visual short-term memory capacity is reached. Our observed patterns of stability are largely consistent with object file theory, with information fading over the course of several fixations after the last fixation of a target object. Irwin (1992a) and Carlson-Radvansky and Irwin (1995) suggested that encoding of identity and position information into object files can be somewhat independent, with some object files containing identity information without a positional index to the scene. We also find evidence for independent encoding patterns for identity and position information. Presence information can be encoded without direct fixation and does not accumulate from multiple fixations. Conversely, to be assimilated, position information requires direct fixation and thereafter shows clear accumulation from multiple fixations of objects. Our observed patterns of encoding for presence and position can also account for the possibility that object files might contain identity information but not position (Carlson-Radvansky & Irwin, 1995; Irwin, 1992a). Presence may be encoded before position because presence information can be encoded without direct fixation, whereas it is necessary to fixate objects to extract position labels. Prior to fixation of a target object, therefore, it is possible that presence information has already been encoded, but position information about the object has yet to be extracted.

While our data are largely consistent with object file theory, two aspects of our data are not. First we find that position information is accumulated from multiple fixations. This is in contrast to the Irwin and Andrew's (1996) suggestion that little or no information is accumulated across fixations. We also find evidence that some object properties may be retained stably and do not fade over the next few fixations following the last fixation of a target object. This result is not consistent with the predictions of object file theory.

Henderson and colleagues (Henderson & Hollingworth, 1999; Henderson et al., 1999; Hollingworth & Henderson, 2002; Hollingworth et al., 2001) have offered a third view of the retention of object information from complex scenes. They argue that object descriptions are not limited to only 3–4 objects at a time; rather many object descriptions can coexist. Hollingworth and Henderson (2002) found that object properties accumulate with increasing fixation time. Furthermore they argue that object descriptions are not as transient as suggested in object file theory but are retained stably after attention is withdrawn from an object. Our observed accumulation of object position information is consistent with these ideas. Similarly, we found that some object properties may be retained stably after attention is withdrawn. However, accumulation and stability of retention were limited, and for most of the object properties we tested our data did not show these patterns.

Of the existing models of the encoding and retention of object descriptions, our data are most consistent with Irwin's object files. However, the current results extend our understanding of their structure and properties in scene viewing.

Object files across multiple fixations

Descriptions of objects can be encoded and retained from visual scenes. These descriptions comprise multiple object properties, including object presence (similar to identity), colour, and position information. However, as the current experiments suggest, not all object properties are encoded and retained.

Retained features show differential patterns of encoding. Object presence and colour show evidence for extrafoveal extraction and do not accumulate from multiple fixations, suggesting that no extra extraction of these properties occurs during fixations of the objects. Conversely, position information appears to require foveation and accumulates from each refixation of the target object. Thus, each time the eye returns to an object, information regarding its position is resampled and accumulated, whereas presence and colour are not. Object descriptions show differential stability of the component properties. Hence we cannot consider object descriptions as unitary constructs, but rather as files in which the component properties are somewhat independent.

Object descriptions in which there is extraction and accumulation of position information from each fixation of an object, but no accumulation of presence and colour information, are behaviourally plausible. In the real world, object properties such as identity and colour are unlikely to change, and so resampling and accumulating these sources of information is not necessary. However, the position of an object is likely to change, particularly for objects that we might manipulate or that may themselves move. For this reason, continued extraction and encoding of position information from each fixation of an object may serve not only to accumulate position information about unmoving objects, but also to update position information if an object has moved within the scene. The observed patterns of extraction therefore allow for a retained description of objects that is dynamic and able to encompass changing positions of objects, while encoding other object properties, which are unlikely to change.

This would suggest that object files can be constructed before the first fixation of an object and contain identity and colour information. Upon fixating the object a position tag becomes associated with the already established object file (for a similar position, see Irwin, 1992a). Our data suggest that position information continues to be sampled, and can therefore accumulate, on each subsequent fixation of the object. This resampling allows for the object file to be updated and so to register any change in the position of the object in the scene. It is also clear from these data that the manner in which the components of object files are encoded and retained depends on the particular property. Transience of identity information along with stability of position or colour information allows for the possibility that object files can lose identity information yet endure, retaining information about position or colour. The persistence of any object file does not appear to depend on the presence of any particular property information; in this way object files are robust to an ever changing visual environment.

REFERENCES

- Antes, J. R., & Penland, J. G. (1981). Picture context effects on eye movement patterns. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 157–170). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, 15(6), 719–722.
- Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye-movements. *Nature*, 371(6497), 511–513.

- Carlson-Radvansky, L. A., & Irwin, D. E. (1995). Memory for structural information across eye-movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1441–1458.
- Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception & Psychophysics*, 62(4), 673–683.
- De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research—Psychologische Forschung*, 52(4), 317–329.
- De Graef, P., Detroy, A., & d'Ydewalle, G. (1992). Local and global contextual constraints on the identification of objects in scenes. *Canadian Journal of Psychology—Revue Canadienne De Psychologie*, 46(3), 489–508.
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36(12), 1812–1837.
- Dodge, R. (1900). Visual perception during eye movement. *Psychological Review*, 7, 454–465.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Atkins (Ed.), *Perception: Vancouver studies in cognitive science* (Vol. 2, pp. 89–110). New York: Oxford University Press.
- Henderson, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General*, 123(4), 410–426.
- Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science*, 8(1), 51–55.
- Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–298). Oxford, UK: Elsevier.
- Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10(5), 438–443.
- Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual-attention and extrafoveal information use during object identification. *Perception & Psychophysics*, 45(3), 196–208.
- Henderson, J. M., & Siefert, A. B. C. (1999). The influence of enantiomorphic transformation on transsaccadic object integration. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 243–255.
- Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 210–228.
- Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception* (pp. 309–331). New York: Holt.
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual-attention in saccadic eye-movements. *Perception & Psychophysics*, 57(6), 787–795.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113–136.
- Hollingworth, A., 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(4), 761–768.
- Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, 6(1), 1–12.
- Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 7(3), 604–610.
- Irwin, D. E. (1991). Information integration across saccadic eye-movements. *Cognitive Psychology*, 23(3), 420–456.
- Irwin, D. E. (1992a). Memory for position and identity across eye-movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 307–317.
- Irwin, D. E. (1992b). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146–165). New York: Springer-Verlag.
- Irwin, D. E. (1993). Perceiving an integrated visual world. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience* (pp. 121–142). Cambridge, MA: MIT Press.
- 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., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye-movements. *Perception & Psychophysics*, 34(1), 49–57.
- Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, 64(6), 882–895.
- Johansson, R. S., Westling, G. R., Backstrom, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *Journal of Neuroscience*, 21(17), 6917–6932.
- Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.
- Lamare, M. (1892). Des mouvements des yeux dans la lecture. *Bulletin et mémoire de la société française d'ophtalmologie*, 10, 354–364.
- Land, M. F. (1993). Eye–head coordination during driving. *IEEE Systems, Man and Cybernetics Conference Proceedings*, Vol. 3, pp. 490–494.
- Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, 369(6483), 742–744.
- Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11), 1311–1328.
- Latour, P. (1962). Visual thresholds during eye movements. *Vision Research*, 2, 261–262.
- Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, 3, 525–551.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Martin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, 81, 899–917.
- McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3), 563–581.
- McConkie, G. W., & Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (pp. 137–162). Newark, NJ: International Reading Association.
- McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, 25(3), 221–224.
- Melcher, D. (2001). Persistence of visual memory for scenes: A medium-term memory may help us to keep track of objects during visual tasks. *Nature*, 412(6845), 401.
- 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.
- O'Regan, J. K. (1992). Solving the real mysteries of visual-perception: The world as an outside memory. *Canadian Journal of Psychology—Revue Canadienne De Psychologie*, 46(3), 461–488.
- 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(1–3), 191–211.
- O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does transsaccadic fusion exist. *Vision Research*, 23(8), 765–768.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(6), 939–973; discussion 973–1031.
- Pollatsek, A., & Rayner, K. (1992). What is integrated across fixations? In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 166–191). New York: Springer Verlag.
- Rayner, K. (1978). Foveal and parafoveal cues in reading. In J. Requin (Ed.), *Attention and performance* (Vol. 7, pp. 149–162). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology*, 12, 206–226.
- Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades. *Perception & Psychophysics*, 34(1), 39–48.
- Rayner, K., & Pollatsek, A. (1992). Eye-movements and scene perception. *Canadian Journal of Psychology—Revue Canadienne De Psychologie*, 46(3), 342–376.
- Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual Cognition*, 7(1–3), 17–42.
- Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Vision Research*, 40(10–12), 1469–1487.

- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1995). Image flicker is as good as saccades in making large scene changes invisible. *Perception*, 24(Suppl.), 26–27.
- 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(5), 368–373.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, 7(1–3), 127–145.
- Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye-movements and spatial attention. *Quarterly Journal of Experimental Psychology*, 38A, 475–491.
- Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, 7(5), 301–305.
- Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, 32(5), 579–592.
- Tatler, B. W., & Wade, N. J. (2003). On nystagmus, saccades and fixations. *Perception*, 32(2), 167–184.
- Wade, N. J., Tatler, B. W., & Heller, D. (2003). Dodge-ing the issue: Dodge, Javal, Hering and the measurement of saccades in eye movement research. *Perception*, 32(7), 793–804.
- Wolfe, J. M. (1996). Post-attentive vision. *Investigative Ophthalmology & Visual Science*, 37(3), 981.

Original manuscript received 29 May 2003

Accepted revision received 30 June 2004

PrEview proof published online 21 October 2004