Intraub, H. (1999). Understanding and remembering briefly glimpsed pictures.
In Fleeting Memories: Cognition of Brief Visual Stimuli (pp. 47-70).
V. Coltheart (Ed.), MIT Press, Cambridge:MA

Chapter 3

Understanding and Remembering Briefly Glimpsed Pictures: Implications for Visual Scanning and Memory

Helene Intraub

The world we perceive is made up of complex, interrelated collections of objects and backgrounds. Yet, at no moment in time does an observer perceive a scene in its entirety. During visual scanning, each eye fixation delivers only a partial view of a continuous environment, and each partial view can be quite brief. For example, when studying a picture of a scene, the viewer's average fixation frequency can be as high as 3 per sec (Yarbus 1967). A classic issue in perception has been how to account for our ability to perceive and remember the meaningful spatial relations that comprise real-world scenes, given the brief and piecemeal character of the visual input.

Research on picture perception and memory provides one avenue for addressing this question. Hochberg (1978, 1986), for example, proposed a theory of visual integration of partial views based primarily upon an analysis of the viewer's ability to understand dynamic pictorial displays. In one experiment he demonstrated that when viewers looked at a moving display through an aperture, although they never saw the whole display at once, they rapidly integrated their partial views into an understandable whole. He proposed that the perception of the whole is achieved through the use of a mental schema—an abstract representation that serves to guide interpretation and integration of successive views of the visual world. The schema can be thought of as an abstract representation of the layout and major landmarks that characterize a scene without preserving sensory detail. Hochberg (1986) argued that, contrary to our subjective experience, "there is much in the world that simply goes unnoticed and unrepresented in the structure that we use to store and to assimilate new views" (22:58).

Consistent with this characterization of scene perception, recent research has shown that observers are surprisingly poor at detecting a change in a scene when the change is preceded by a transient such as a brief mask (Rensink, O'Regan, & Clark 1997) or a shift in viewpoint (Simons & Levin 1997). For example, in one experiment, a videotape was made of two actors at a table, talking. When the camera panned away from the table to focus on one actor for 4 sec, a central item on the table (a large soda bottle that had been conspicuously used) was replaced with a cardboard box. When the camera returned to the original view, remaining for 30 sec, none of the viewers noticed that the bottle had "become" a box (Simons 1996).

Similarly, it has been demonstrated that various types of changes made to a scene during a saccade (when the eye rapidly travels from one region of a scene to another) have been surprising resistant to detection (Grimes, 1996; McConkie & Currie, 1996; also see O'Regan, 1992). There has been a growing number of studies supporting the idea that transsaccadic memory is abstract rather than sensory in nature (e.g., Irwin, Brown, & Sun 1988; Irwin 1991; McConkie & Zola 1979; Rayner & Pollatstek 1992; see also Intraub 1997).

If a schematic representation of spatial layout underlies the integration of successive views during visual scanning, then each briefly viewed glimpse of the world would have to be very rapidly analyzed and understood, readily activating expectations about the surrounding area. This possibility has been explored in various ways. In this chapter, I will describe three lines of research from my own lab that are relevant to this possibility. These are (a) recognition memory for briefly glimpsed pictures, (b) target detection during rapid continuous presentation of pictures, and (c) memory for a scene's boundaries, that is, its spatial expanse.

Memory for Briefly Glimpsed Pictures

When photographs are presented in rapid succession, at rates that mimic or surpass the average fixation frequency of the eye, the viewer's ability to recognize them moments later is very poor. At a rate of 3 pictures per sec (similar to the scanning rate reported by Yarbus 1967), recognition memory performance on tests using dissimilar distractors has shown that subjects remember fewer than 50% of the stimuli (e.g., Potter 1976; Potter & Levy 1969). Why is recognition memory so poor following a presentation rate that mimics the rate at which scenes are normally fixated?

One possibility is that under these artificial conditions of presentation, most of the unrelated scenes that flash by are not understood by the viewer and are therefore not remembered. After all, although the rapid sequences used in these experiments mimic the rate at which viewers can shift their fixation during visual scanning, they do not mimic the redundancy and continuity of the information gleaned from successive fixations on a stable environment. The continuity among successive views during normal scanning may result in the development of expectations that serve to guide and facilitate perception. Therefore, although the scanning rate is quite rapid, the viewer has a good idea of what each fixation will bring into view (Neisser 1976). On the other hand, poor memory following rapid continuous presentation of unrelated pictures might be caused by the viewer's inability to identify rapidly changing, unpredictable visual events. However, research on scene perception has suggested that the meaning and layout of a scene are acquired so rapidly that they can affect identification of objects within the scene, at durations as brief as 50–150 msec (Biederman 1972, 1981; Biederman, Mezzanotte, & Rabinowitz 1982). Thus, poor memory following rapid presentation may actually reflect limitations on memory rather than on perception.

Figure 3.1
Examples of the pictures of objects that were used in the brief presentation and RSVP experiments The actual stimuli were presented in color. (Intraub 1979, 1980, 1984, 1985, 1989.)

To explore the viewers' ability to remember briefly presented unrelated pictures, I conducted a series of experiments that tested recognition memory under a variety of time constraints. This research showed that recognition memory for briefly glimpsed pictures could be quite good, as long as viewers were allowed some time between pictorial presentations. This strongly suggested that the poor performance following rapid continuous presentation was not attributable to the viewer's inability to extract information from a very brief glimpse of an isolated pictorial event. Figure 3.1 provides a sample of photographs of objects used in these experiments.

In one series of experiments, Intraub (1979) presented 12 color photographs of common objects for 110 msec each at 6 different stimulus onset asynchronies (SOAs). These ranged from 110 msec to 1.5 sec. At the shortest SOA (110 msec) there was no interstimulus interval (ISI); each briefly presented picture was immediately followed by the next. At the slower SOAs, each 110-msec picture was followed by a gray field that remained on the screen for the full ISI. To test the effects of "nameability" on picture memory, photographs of objects were selected that were "equally easy to see," based upon visual duration thresholds, but that were not equally easy to name, based on each picture's mean naming latency. Subjects' memory was tested using either free recall or a serial recognition test in which 12 nonconfusable objects served as distractors.

Recognition memory for the briefly glimpsed pictures improved markedly from 19% to 92% (corrected for guessing) as the time between stimuli was increased. Free recall of the pictures yielded the same pattern of results, with percent recalled ranging from 13% in the fastest presentation condition to 52% in the slowest. The possibility that this improvement was due to verbal mediation at slower rates (see Paivio 1971; Paivio & Csapo 1969) was not supported because there was no relationship between a picture's nameability (in terms of its naming latency) and whether or not it was remembered. This was true both for recognition and for recall. It seemed likely that something other than a simple name was being encoded during the brief ISIs that were tested—perhaps information related to pictorial detail. This possibility was supported when it was shown that the ability to remember left-right orientation of objects presented for 110 msec increased with increases in ISI (Intraub 1980).

Recall, recognition, and memory for mirror reversal all improved when the time between briefly presented pictures was increased. Apparently encoding continued following offset of a complex visual event. These early experiments, however, could not determine the degree to which this poststimulus encoding may have relied on the presence of an iconic representation of each picture. The gray field shown during the ISI would not be expected to act as a visual mask.

If during continuous presentation (no ISI) visual masking of each picture by the next was the primary reason for poor recognition performance, then in a sequence with ISIs, presentation of a visual mask rather than a blank field during the ISI should cause a dramatic decrease in recognition memory. To test this prediction, Intraub (1980) compared recognition memory for photographs of objects under conditions in which the ISI did or did not contain a mask (see figure 3.2).

One hundred and fifty color photographs were presented for 110 msec each, at a rate of 1 picture every 6 sec (see diagram at the top of figure 3.2). In one condition a blank field was presented during the ISI. In another condition, to present a display during the ISI that would have the same masking potential as pictures in a rapid sequence, two pictures from the same stimulus pool were selected. Each picture was used for half of the subjects in this condition. The subjects were familiarized with their designated picture prior to presentation, and then that picture was presented during each ISI in the sequence. Therefore, although subjects in the two ISI conditions were allowed the same amount of time between stimuli, in the blank-field ISI condition no masking stimulus was presented, whereas in the repeating-picture ISI condition each stimulus was followed by a potential visual mask. Finally, a third group of subjects viewed the same 150 pictures for the same 110 msec duration each, but with no ISI (this is depicted in the lower diagram in figure 3.2). In all three conditions, subjects were instructed to pay attention to and to try to remember the briefly presented pictures. If the poor performance typically obtained following rapid continuous presentation is caused by visual masking of each picture's iconic representation, then memory in the repeating-picture ISI condition should drop to this same low level.

After viewing 150 unrelated pictures, subjects in the blank-ISI and repeating-ISI conditions recognized 77% and 73% of the test items, respectively. The 4% difference in performance did not approach significance. Other experiments showed that presentation of a picture during the ISI did not diminish recognition memory for left-right orientation (Intraub 1980; experiment 3), nor did it diminish the viewer's ability to report pictorial details (Intraub 1980, experiment 4). However, when the same briefly presented pictures were shown with no ISI, recognition accuracy plummeted to 21% correct.

It is important to point out that across experiments, the small decrement in recognition memory between the blank-ISI condition and the repeating-ISI condition sometimes reached significance and sometimes did not (Intraub 1980, 1984). This small difference may have been due to the disruptive effects of visual masking on the iconic representation of each picture (see Loftus, Hanna, & Lester 1988; Loftus, Johnson, &

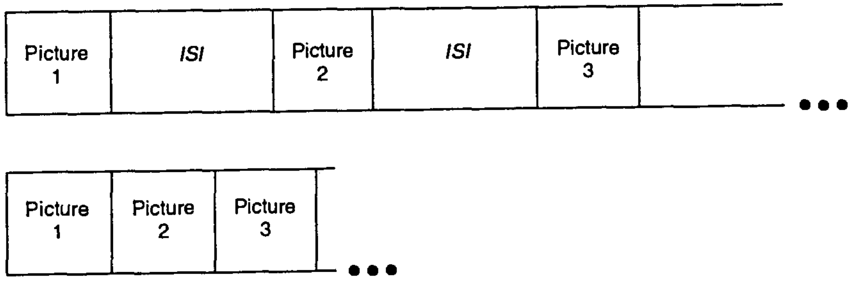

Figure 3.2
Schematic representation of the sequences when there was an ISI (top) and when there was no ISI (bottom). Stimulus pictures were presented for 110 msec each in all conditions. Depending on condition, either a blank field or a familiar picture was presented during the 5890 msec ISI. (Intraub 1980, experiment 1.)

Shimamura 1985; Loftus & Ginn 1984). However, this relatively small decrement in performance does not approach the level of poor performance obtained in the continuous presentation condition. This suggested that something in addition to visual masking was contributing to the viewers' inability to remember pictures that they had just viewed.

Visual Search, Memory, and the Conceptual Short-Term Store

Some years earlier, Potter (1976; also see Potter 1993) had proposed that during rapid continuous presentation, unrelated pictures are momentarily understood and then immediately forgotten. She argued that although each picture is momentarily grasped, conceptual processing initiated by the next *new* picture in the sequence disrupts consolidation in memory.

Potter argued that a picture is identified within about 100 msec. Until the item is identified, it is vulnerable to visual masking by a new visual event. Once identification is complete, however, the pictorial representation is maintained in a conceptual short-term store for a few hundred milliseconds. It is during this time that the information is consolidated in memory. Although the representation is no longer subject to visual masking, its consolidation into memory can be disrupted by "conceptual masking." Conceptual masking can occur if a new, *meaningful* visual event is presented to the observer before consolidation of the previous event is complete. The new, meaningful visual event elicits conceptual processing and replaces the previous item in the conceptual short-term store. Potter argued that the store can maintain only one item at a time. If a picture is consolidated in memory before onset of a new picture, it is likely to be remembered.

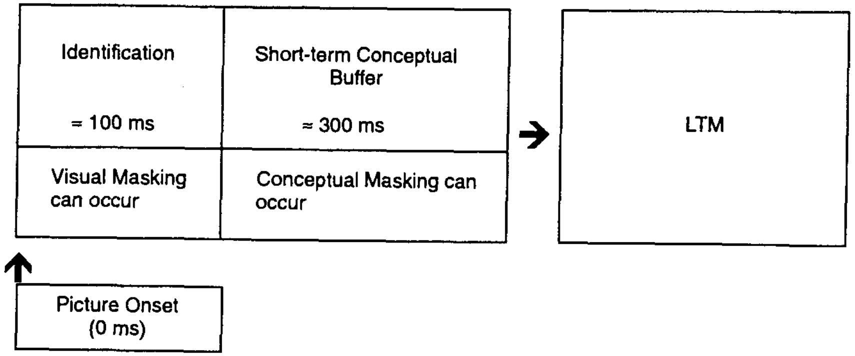

Figure 3.3
Schematic representation of Potter's (1976) model.

If not, although the viewer will have momentarily understood the picture, it will be forgotten by the time he or she has understood the next picture in the sequence. A schematic diagram of Potter's model is presented in figure 3.3.

The relatively good recognition memory for briefly glimpsed pictures that is obtained when a second or more is allowed between presentations (even when a repeating picture is presented during those ISIs) is consistent with this proposal, because the timing would eliminate the interfering effects of a conceptual mask.

To test if subjects could indeed momentarily grasp more than they could remember during continuous presentation, Potter (1976) contrasted subjects' ability to detect cued pictures with other subjects' ability to remember pictures at the end of each sequence. She reasoned that if viewers understand more than they can remember, on-line detection will be superior to immediate recognition memory. Conversely, if memory is poor following rapid presentation because observers can identify only a few unrelated visual events and remember only those that were identified, then detection accuracy should be no better than recognition memory.

She presented 16 unrelated color photographs of scenes in a continuous stream (no ISIs) at SOAs ranging from 113 to 333 msec per picture. One picture in each sequence was cued either by being presented in advance or by being described in advance, using a brief verbal title (e.g., "a road with cars"). The rationale was that whereas a correct detection based on the pictorial cue could be made on the basis of physical characteristics alone, correct detection based on the verbal cue could occur only if the viewer conceptually identified the target picture's meaning as it flashed by. Comparison of the two conditions was intended to provide insight into the extent to which target detection relies on specific visual detail as opposed to conceptual analysis. The proportion of targets detected based on the verbal cue was interpreted as reflecting the minimum proportion of pictures identified during presentation of the sequence. In the recognition memory condition, subjects were instructed to remember as many pictures as possible, and immediately following each sequence they participated in a recognition test in which all 16 presentation pictures were mixed with 16 distractor pictures.

In the detection conditions, subjects pressed a key as soon as they detected the cued picture. Responses falling between 250 and 900 msec following target onset were counted as correct. Results showed that subjects could identify unrelated scenes remarkably well at presentation rates that were equivalent to the average eye fixation frequency of 3 per sec, and those that were considerably more rapid. On the basis of the verbal cue, detection accuracy for rates of 113, 167, 250, and 333 msec/picture was 64%, 74%, 89%, and 78%, respectively (the apparent decrease at the slowest rate was due to an increased number of anticipation responses). Performance with the verbal cue was almost as good as seeing the picture itself in advance. Furthermore, in all cases, detection accuracy far surpassed recognition memory, with recognition accuracy ranging from 11% correct at the fastest rate to 42% correct at the slowest rate.

These results showed that completely unrelated scenes presented at rates equal to or faster than the average fixation frequency could be conceptually analyzed and matched to a verbal cue, although memory for the pictures immediately following presentation was relatively poor. The results were replicated and extended by Intraub (1981), using a number of design modifications that provided perhaps even more compelling support for Potter's (1976) model.

Although Potter (1976) had used general titles as cues to avoid facilitating perception of the target picture, one could argue that "a road with cars" narrows the visual expectations of the viewer enough to enhance perception of the target picture (see, e.g., Carr & Bacharach 1976; Neisser 1976). This would artificially inflate the detection rate. Intraub (1981) attempted to reduce expectancy by using not only descriptive titles as cues in a search task, but also cues that provide no specific visual details about the target. These latter cues were "category cues" and "negative cues." This was accomplished by presenting subjects with 12-item sequences that contained 11 items from a single general category (e.g., animals) and 1 from a different category. The odd item was cued by its basic-level name (e.g., chair), its superordinate category (e.g., "house furnishings or decorations"), or a negative cue (e.g., the item that is not an animal). The pictures used in this experiment were color photographs of objects similar to those shown in figure 3.1.

Sequences were presented at rates of 114, 172, and 258 msec/picture with no ISIs. Photographs were selected that were as visually dissimilar as possible. For example, the category "animals" contained creatures as diverse as a frog, a dog, a giraffe, and a butterfly. The target picture did not differ noticeably in size or general coloration from the other pictures in the sequence. The subject's task was to press a key upon detection, and then to describe the target picture. This meant that unlike Potter's experiment, rather than inferring correct detection from the key press alone, correct detection was based on the subject's ability to describe the target. Finally, to provide subjects in the recognition condition with a better chance to demonstrate good recognition memory, Intraub (1981) used a less taxing recognition test than Potter had (4 items instead of 32). The test always included two pictures that had been shown during presentation (one was the "odd" picture) and two distractors (one from the general category presented in the sequence, and one from a new category that differed from all the pictures, including the "odd" picture).

The results were consistent with those obtained by Potter (1976). As can be seen in table 3.1, subjects were able to detect and describe target pictures whether they had been cued by name, by superordinate category, or by negative cue. Furthermore, in all conditions, detection accuracy surpassed immediate recognition memory. The viewers' success in detecting and describing these target pictures on the basis of cues that provided no direct information about their visual characteristics, strongly supports the argument that a large proportion of the rapidly presented pictures were at least momentarily understood although they were rapidly forgotten. It is interesting to note that when subjects in the recognition condition were asked to describe the sequences they had just seen, all reported that the pictures were grouped by category and 83% specifically reported that there was a category-plus-odd-picture arrangement. In spite of having noticed this, they were very poor at recognizing the odd picture immediately following presentation.

The detection experiments showed that viewers can very rapidly identify a picture's meaning, but that identification alone is not sufficient to ensure later recognition. What is it that prevents consolidation during rapid presentation? To test Potter's (1976) argument that it is the presentation of a new, meaningful visual event (i.e., a conceptual mask), Intraub (1984) presented 16 pictures for 112 msec each with a 1.5 sec ISI. The ISI contained a blank field, a repeating to-be-ignored picture, or a new picture each time (see upper diagram in figure 3.2). Subjects were instructed to memorize the briefly presented stimulus

Table 3.1
Proportion of pictures detected by name, category, and negative category, and the proportion of target pictures recognized at each rate of presentation.

Rate (msec/picture)	Detection			Recognition Memory*
	Name	Cat.	Neg.	Target
258	.89	.69	.79	.58
172	.86	.71	.58	.49
114	.71	.46	.35	.19

Cat., category; neg., negative category.
*Recognition scores were corrected for guessing.
Based on Intraub 1981, table 1.

pictures. Once again, there was not a large difference in performance between the blank-ISI and the repeating-ISI conditions: recognition accuracy was 89% and 80% (which did not differ significantly). However, when a new, to-be-ignored picture was presented in the ISI each time (changing-ISI condition), recognition accuracy for the briefly presented pictures dropped to 63% correct.[1]

What is it about a new, to-be-ignored picture that disrupts processing more than a familiar repeating picture? According to Potter's view it is the conceptual processing that is elicited by each new, meaningful picture. However, another possibility is that it is not the "meaningfulness" of the new ISI pictures per se, but the novelty of each new ISI picture, that draws the subject's attention. To tease apart these two possible forms of interference, Intraub (1984) presented one of the following during each ISI: (a) a familiar, repeating picture, (b) a new picture, or (c) a new, colorful nonsense picture. The nonsense pictures were created by tracing the outline of each changing-ISI picture and filling it with splotches of color that did not follow any of the original contours within the picture. A new, nonsense picture was no more disruptive to memory than was a familiar, repeating picture: subjects recognized 87% of the pictures in the changing-nonsense-picture condition, and 90% in the repeating condition. Presentation of a new picture during each ISI, however, resulted in a significant reduction in recognition accuracy to 73% correct. During rapid presentation, apparently it is not simply the presentation of a new visual event that disrupts consolidation of individual pictures, but presentation a new *meaningful* event—a conceptual mask.

Loftus and Ginn (1984) provided converging evidence for conceptual masking and demonstrated a clear distinction between visual masking

and conceptual masking. Pictures were presented for 50 msec, followed by either a visual noise mask or a conceptual mask (a new meaningful picture). The mask occurred either immediately or after a 300 msec delay. In addition, the mask was presented at one of two illumination levels. As predicted, in the immediate condition, luminance (a variable known to affect visual masking) affected the subjects' ability to report details of the picture, but type of mask did not. At the 300 msec delay (which allowed enough time for identification to be complete), luminance did not affect report, but type of mask did. A display with new conceptual information (i.e., a picture mask) resulted in fewer correct details being reported than did the visual noise mask.

In Intraub (1984), although presentation of a new picture in the ISI disrupted encoding more than had a familiar, repeating picture, memory did not drop to the low level obtained following rapid continuous presentation (19–21% correct),[1] a condition in which numerous new pictures follow a given picture. This argues against Potter's (1976) claim that the conceptual short-term store can hold only one picture at a time and that the information is "bumped" from the buffer when conceptual processes are elicited by the *next new picture in the sequence.* Subjects apparently could continue some processing of the briefly glimpsed pictures during the changing-ISI sequence. Given this result, Intraub (1984) sought to determine if the ability to continue processing was to some degree under the observer's control.

Subjects viewed the changing-ISI condition but were now provided with one of three attention instructions: attend to the brief pictures (112 msec), attend to the long-duration pictures (1.5 sec), or attend to both types of pictures equally (see upper diagram of figure 3.2). If conceptual masking by the new ISI picture automatically "bumps" the previous briefly presented picture from the store, then attention instructions should have no effect on memory. Subjects' recognition accuracy for the long and brief pictures as a function of attention instruction is shown in figure 3.4. Results showed a pronounced effect of attention instructions on memory for both the 112 msec and the 1.5 sec pictures. This demonstrated that the viewer has more control over these early stages of processing than was originally thought.

Intraub (1984) argued that the conceptual buffer must be able to hold more than one picture at a time. Consistent with this view, Intraub and Nicklos (1981; Intraub 1985a) presented 24 pictures in a sequence in which a blank interval was presented after each picture or after each group of 2, 3, or 4 pictures. The effect of grouping these pictures indicated that the buffer could hold up to three complex pictures. Subjects could encode successively presented pictures as long as the series was not too long.

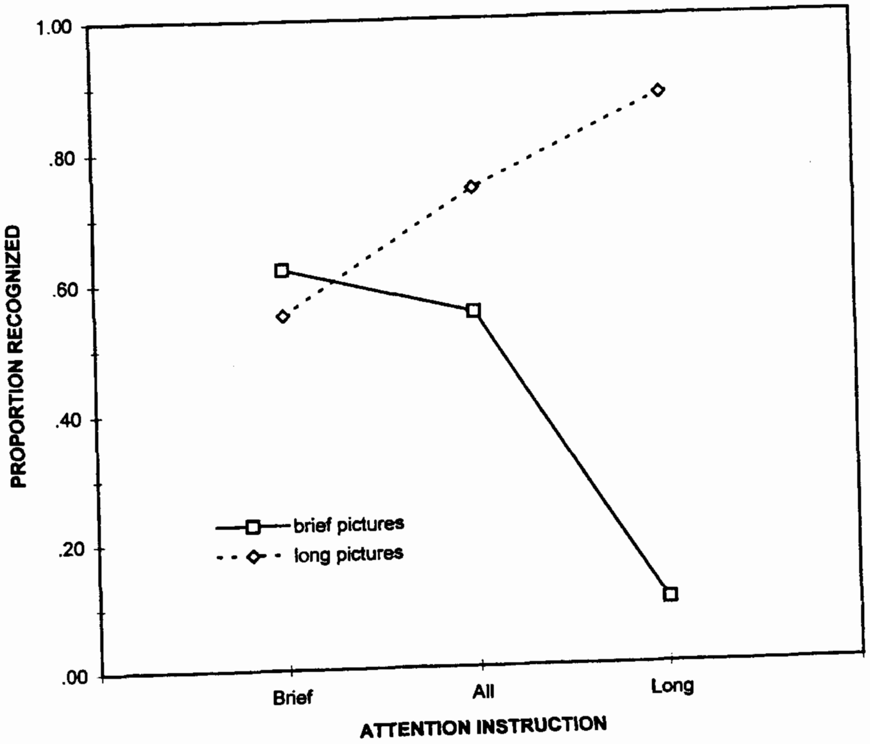

Figure 3.4
Proportion of long-duration and brief-duration pictures recognized (corrected for guessing) following each attention instruction. (From Intraub 1984.)

Visual Dissociation

The conceptual short-term buffer can hold more than one picture at a time. A phenomenon referred to as "visual dissociation" of pictures suggests that identification and integration of components of pictures can be ongoing for more than one picture at a time (Bishop & Intraub 1996; Intraub 1985b, 1989). Visual dissociation occurs during rapid presentation, when a searched-for feature migrates to a temporally adjacent picture in the sequence and the subject perceives an illusory conjunction of the picture and the feature.

In one experiment, Intraub (1985b) presented subjects with color photographs of objects at a rate of 9 pictures/sec. They were instructed to report which object in each 12-picture sequence was surrounded by a black outline frame. Although subjects were often confident and correct, they were frequently confident and wrong (30–50% of the time). Subjects, when wrong, almost always reported the immediately preceding or immediately following object in the sequence as the one with the frame. When subjects made an error, they often described the target object (the one actually in the black frame) as a "frameless" picture. This type of report shows that the subjects had perceived and remembered both the target picture and the temporally adjacent picture. Apparently, both were identified and understood, but under time pressure the frame had become integrated with the wrong stimulus, while both were being simultaneously processed.

In another experiment in which the procedure was the same except that in addition to the verbal response, subjects pressed a key as soon as they detected the frame, reaction time was clearly associated with the "direction" of the frame's migration (i.e., to the preceding or following

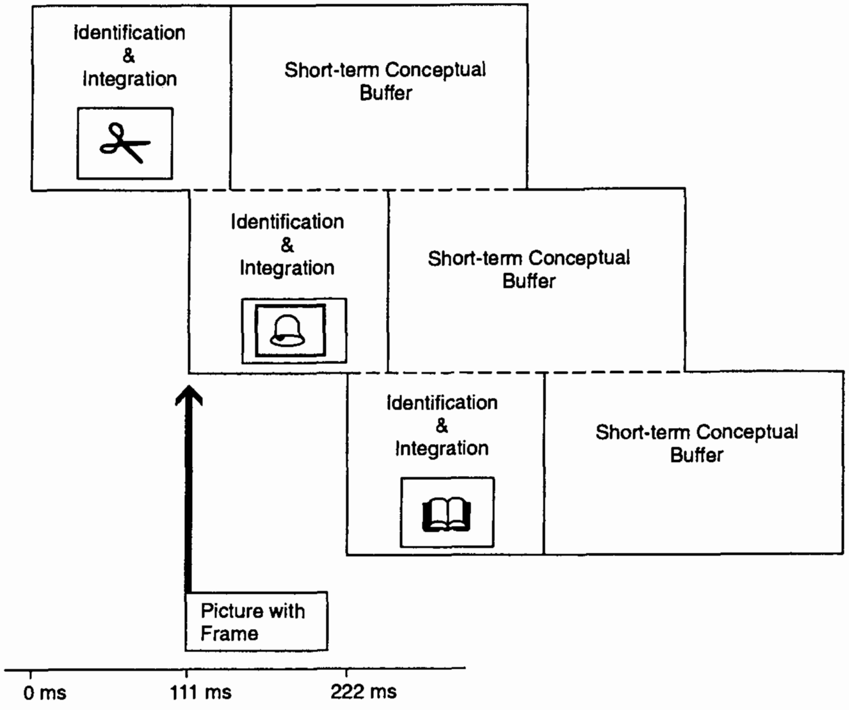

Figure 3.5
In this example, 3 successive items from an RSVP sequence are depicted (actual stimuli were color photographs). The picture of the bell is presented with the black frame. When the frame is detected early, it is likely to become integrated with the previous picture in the sequence (the scissors). When frame detection is relatively late (e.g., an additional 26 msec), it is likely to become integrated with the following picture in the sequence (the book). Dotted lines indicate places of potential overlap, where 2 pictures may undergo similar processing at the same time.

picture in the sequence). Frame detection times were 26 msec faster on those trials associated with frame migration to the preceding picture than those associated with frame migration to the following picture. In other words, when the frame was rapidly detected, it was more likely to become integrated with the "tail-end" of processing associated with the previous picture, but when frame detection was relatively slow, it was more likely to become integrated with processes that were just initiated on a new picture (Intraub 1989, experiment 2). A schematic representation of a modified version of the original Potter (1976) model that characterizes these results is presented in figure 3.5.

The same type of temporal migration was obtained when subjects were required to search for a specified object (e.g., an outline drawing of a chair) in a sequence of outline scenes (e.g., a living room, a bedroom, and a study; Intraub 1989). In this case subjects would frequently, and confidently, report the object as occurring in the preceding or following scene in the sequence. Yet if these stimuli are presented at a rate of 3–4 pictures per sec, the searched-for element (frame or object) does not become integrated with temporally adjacent pictures. At rates such as these, the successive views have apparently reached a stable state in which they are immune to integration errors.

This can also be seen in Hochberg's (1986) aperture viewing experiment. In this case, subjects were informed that the ambiguous display they would see was actually an outline cross moving behind a circular aperture. They could perceive the display as such when the presentation rate was 2–3 views per sec. At more rapid rates of presentation (e.g., 10 views per sec), the views became visually integrated and could not each be evaluated in terms of the mental schema.

The discussion of timing is an important one. Potter (1976) raised the question of why, if the visual system is capable of identifying a picture within 100 msec, visual scanning would have an upper limit of about 3–4 fixations per sec. She speculated that the average fixation frequency of 3 per sec may be a compromise between the need for rapid identification of each fixated part of the visual world (which can occur within 100 msec) and the need to retain some portion of what has been seen (by allowing consolidation of information in the conceptual short-term store). The visual dissociation experiments and the research on the effects of presentation rate on aperture viewing suggest that this limitation on the scan rate also serves the important purpose of minimizing the likelihood of visual integration errors during visual scanning.

To summarize, research on detection and memory of briefly presented views shows that a picture's conceptual content can be very rapidly accessed, but that consolidation in memory takes a somewhat longer amount of time. Although the characteristics of a display can affect consolidation of the previously glimpsed picture, the viewer possesses considerable control over the allocation of attention to previous and current visual events. These capabilities would be instrumental to a system that interprets the visual world by integrating successive views within a larger schematic context. The next question to be considered is whether there is any evidence to suggest that a mental schema is in fact activated by a single glimpse of a scene.

Boundary Extension

If activation of a mental schema occurs in response to the detection of a partial view of a scene, might there be a detectable "residue" of this activation when the scene is later remembered? A possible affirmative answer is provided by research on "boundary extension," a memory distortion reported by Intraub and Richardson (1989). What they observed was that viewers tended to remember having seen a greater expanse of a scene than had been shown in a photograph (see figure 3.6). Subjects' drawings of remembered photographs, and their responses in a recognition/rating test, revealed that they remembered having seen information that did not exist in the photograph, but that would be expected to exist just outside the camera's field of view.

Boundary extension has been observed under a wide variety of conditions in which pictures are presented for relatively long periods of time (e.g., 15 sec). It occurs regardless of whether objects are cropped by the edges of the picture (Intraub, Bender, & Mangels 1992; Intraub & Bodamer 1993); thus ruling out an explanation based on the Gestalt principle of object completion (see Ellis 1955). It occurs following presentation of as few as 3 pictures (Intraub & Bodamer 1993, see demonstration condition) or presentation of as many as 20 pictures for 15 sec each (Intraub & Richardson 1989). It also has been obtained using a variety of memory tests including recall (subjects draw remembered pictures), boundary recognition tests, and even a test in which subjects physically moved markers to indicate the remembered boundaries of a photograph (Intraub 1992; Intraub & Bodamer 1993; Intraub & Richardson 1989; Legault & Standing 1992; Nyström 1993).

A possible explanation of boundary extension is that when viewers detect a partial view of a scene, this activates a mental schema that represents the general layout of what would be seen if the viewer actually could make an eye fixation just beyond the picture's boundaries. In terms of this formulation, comprehension of a partial view involves mentally locating the view with respect to the larger scene. For example, when viewing a portrait-style photograph of a smiling

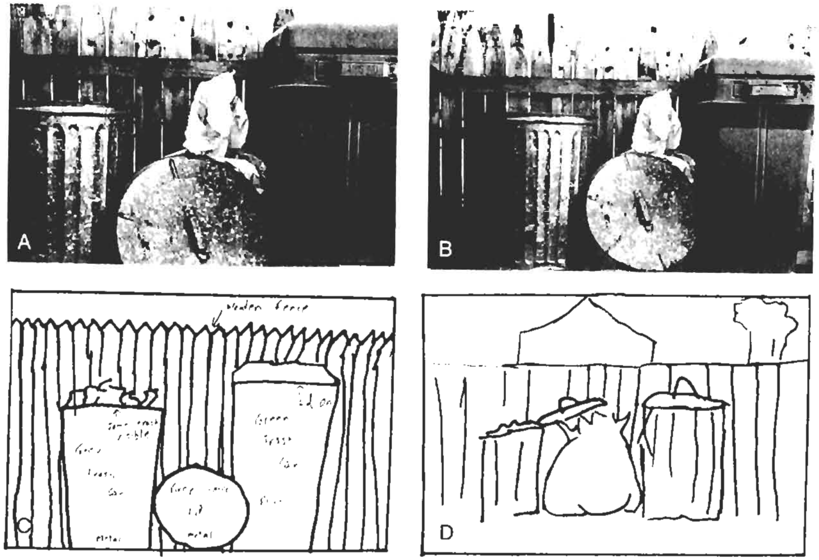

Figure 3.6
Panel A shows a close-up version of a scene with main objects cropped. Panel B shows a slightly more wide-angle version of the scene with no objects cropped. The subjects' drawings in panels C and D are representative drawings of the photographs in panels A and B, respectively. In comparing drawings with the photographs, be sure to pay attention to four picture boundaries. (Intraub & Richardson 1989.)

friend, we do not perceive the picture as depicting a disembodied head. The rest of the friend and the rest of the background are understood to "exist" just outside the picture's boundaries. Interpretation of the partial view includes these spatial assumptions about the layout of the continuous scene from which the partial view was culled.

We have proposed that the viewer remembers seeing not only what was actually presented but also what was understood to exist just beyond the picture's boundaries, thus causing the viewer to recall having seen a greater expanse of the scene than was shown (Intraub 1992; Intraub, Bender, & Mangels 1992). An interesting point to consider is that although this results in a nonveridical representation of the picture, it will often result in a veridical representation of the real-world scene that extends beyond the picture's boundaries. This can be seen in figure 3.6. A comparison of the close-up photograph and the subject's drawing of that photograph shows that boundary extension occurred. If one then compares that same drawing with the wide-angle photograph of the scene, it is clear that although the subject did not correctly remember the close-up, the subject's error included information that actually did exist outside the close-up's boundaries! The "error" shows that the observer had successfully extrapolated information regarding the scene's layout.

Several studies following the initial report of boundary extension have supported the hypothesis that the phenomenon reflects activation of a scene schema during perception. This has been referred to as the "perceptual schema hypothesis." In one test of the hypothesis, Intraub and Bodamer (1993) sought to eliminate the phenomenon by drawing the viewer's attention to the pictures' boundaries. Their rationale was that if boundary extension reflects a fundamental aspect of picture perception (i.e., understanding the partial view within its larger expected context), then it should be difficult to prevent even when the viewer attempts to do so.

Subjects viewed 12 photographs of scenes for 15 sec each and were told to remember them in as much detail as possible, including the background and the layout of the objects in the picture space. In the control condition, that is all they were told. In the test-informed condition, they were told about the boundary memory test in advance—so they knew that memory for the pictures' boundaries would be especially important. In the demo condition, prior to the experimental sequence, they were given a demonstration of boundary extension using 3 pictures. All subjects drew pictures with extended boundaries. This was pointed out to them, and the phenomenon was described. Prior to presentation of the experimental sequence, they were specifically told to try to prevent the distortion from occurring.

Both a drawing task and a recognition memory test (in which distractors showed both boundary extension and boundary restriction) revealed that subjects in the test-informed and demo conditions experienced boundary extension. The effect was attenuated in comparison with the control condition, but was not eliminated. Even when subjects were forewarned, they were unsuccessful in their attempts to prevent its occurrence.

In other research, Intraub, Gottesman, and Bills (1998) directly tested the perceptual schema hypothesis by comparing memory for spatial expanse in a condition in which the pictures depicted a partial view of a scene (which should activate the schema), and a condition in which the pictures did not (and therefore should not active the schema). They presented subjects with 16 photographs of scenes, 16 outline drawings of the scenes (traced from the photographs), or 16 outline objects (traced from the photographs without including any indication of a background). Both a boundary recognition test and a drawing task showed that boundary extension occurred for both types of scenes, but not for the pictures of objects that had been presented without scene structure. Converging results were reported by Legault and Standing (1992), who found that photographs of objects in scenes yielded boundary extension, whereas pictures showing an outline drawing of the main object from each scene did not.

The type of schematic expectations that yield boundary extension are apparently limited to cases in which the viewer understands the stimulus to be a partial view of a continuous scene. To determine if the perceptual schema could be activated via a top-down route, Intraub et al. (1998) presented the outline objects, described the photographs from which the objects had been traced, and instructed the subjects to attempt to imagine the photograph and "mentally project" it onto the outline object. Subjects were instructed to remember the size and placement of each object in as much detail as possible. They were told that the purpose of the exercise was to determine if imagining the scene from which the object was drawn would aid memory.

While subjects viewed each of the 16 outline objects for 15 sec each, the experimenter read a description of the photograph that the subject was to try to imagine. For example, while studying the outline drawing of a man in a cross-legged sitting position, the subjects heard the following, "The man is sitting on a concrete pebbled ground. Directly behind him is a red brick wall which fills the picture space." In the control condition, subjects were also instructed to remember the size and placement of each object in as much detail as possible, but no mention of scenes was made.

Although subjects in both conditions saw the same stimuli (outline drawings of objects), subjects remembered the size of the objects differently in the two conditions. In the control group, there was no overall directional distortion. Some objects were remembered as being larger and some as being smaller[2]. In the imagination group, however, results yielded the typical pattern associated with boundary extension. Consistent with remembering a more wide-angle view, subjects tended to remember the main object as having taken up less of the picture space. Furthermore, as has been found repeatedly with photographs of scenes, objects traced from close-up versions of the pictures showed a greater degree of distortion than did objects traced from the wide-angle views.

Finally, to determine if it was the imagination task per se that resulted in boundary extension, they again presented the outline objects along with an imagination task. However, this time, the subjects were provided with a description of each object's colors and were asked to mentally project those colors onto the outline objects (with no mention of scenes or partial views). In this case, boundary extension was not obtained and the pattern of results was the same as in the control condition. Apparently, the schema is only activated in response to the understanding of a partial view of a potentially continuous environment (i.e., a scene).

This research suggests that activation of the perceptual schema can be initiated by either bottom-up or top-down indications of a continuous scene. Perception and imagination of scenes apparently share a common underlying structure, a notion that has been applied to other types of stimuli (e.g., Farah 1988; Finke 1985; Reisberg 1996). For the purposes of the present discussion, however, the most important point is that when the observer encounters a partial view of a continuous scene (regardless of whether it is physically presented or imagined), expectations about the continuation of the scene's layout are automatically activated. If the same schema also underlies the integration of eye fixations, then we should be able to find evidence for it under conditions that mimic the temporal characteristics of visual scanning.

Intraub, Gottesman, Willey, and Zuk (1996) conducted two experiments using brief presentations of pictures to address this question. In experiment 1, they sought to determine if boundary extension would occur following a brief exposure to a scene. Prior to their experiments, multisecond stimulus durations had been used in boundary extension experiments, with durations usually as long as 15 sec. If boundary extension reflects the activation of a schema that serves to guide and integrate successive views, then a brief glimpse should be sufficient to cause boundary extension. On the other hand, if a single glimpse allows only for the extraction of the "gist" of a scene (in terms of its semantic meaning alone), then boundary extension should not occur. In the latter case, although subjects would not be expected to have an accurate memory for the boundaries (i.e., they might sometimes expand them and sometimes contract them), without schema activation there would be no reason to expect a unidirectional error.

To test these hypotheses, Intraub et. al (1996) used a design that was similar in many ways to the recognition memory experiments described earlier (e.g., Intraub 1979, 1980), except that memory for boundaries was tested. Seven scenes were presented with an SOA of 5 sec. The pictures were presented for either 250 msec each or 4 sec each, with a black-and-white visual noise mask during the ISI. Subjects saw either close-up or wide-angle views of the same 7 scenes. Presentation was followed by a drawing task and a boundary recognition task. Subjects' performance on both types of tests showed that they remembered the pictures with extended boundaries. If anything, the unidirectional distortion was slightly more pronounced in the 250 msec condition than the 4 sec condition. Memory for pictures presented at both stimulus durations replicated the pattern of results reported in numerous boundary extension experiments (e.g., Intraub et al. 1992; Intraub & Berkowits 1996): close-ups yielded a strong degree of boundary extension and wide-angle views yielded little or no directional distortion.

This experiment clearly demonstrated that boundary extension is not limited to stimuli with long durations. To explore the early time course of the phenomenon further, in another experiment Intraub et al. (1996) sought to determine if the same memory distortion would occur with an SOA that is brief enough to mimic a fixation frequency of 3 views per sec. Subjects were presented with 72 sequences. Each contained 3 unrelated photographs of scenes presented for 333 msec each with no ISI. The third picture was followed by a visual noise mask for 1 sec, followed by the test picture. The test picture was always identical to one of the pictures the subject had just seen. Subjects rated the view depicted in the test picture using a 5-point boundary scale (ranging from a score of -2 to +2) as "much too close," "slightly too close," "same," "slightly too far," or "much too far," and then provided a confidence rating of "sure," "pretty sure," or "not sure."

What happened was that test pictures tended to be rated as being "too close" to be the original, which indicates that the original was remembered with extended boundaries. The mean boundary scores and .95 confidence limits as a function of the serial position of the picture that was tested, are presented in table 3.2. As can be seen in the table, the directional bias was significant regardless of the serial position of the stimulus, and the removal of "not sure" responses, if anything, resulted in a stronger unidirectional distortion of the pictures' boundaries. Given only 3 pictures presented in rapid succession, and a 1 sec retention interval, boundary extension occurred. In addition to grasping the conceptual meaning of each unrelated, briefly glimpsed picture, viewers had apparently extrapolated each scene's layout. Merely a glimpse of a scene was sufficient to activate schematic expectations beyond the picture's physical boundaries.

Summary and Conclusions

We possess a fleeting memory for unrelated objects and scenes that are presented in rapid succession at rates that mimic or surpass the average fixation frequency of the eye. At presentation rates ranging from about 110 msec to 333 msec per picture, the subjective experience is a very interesting one of grasping and losing large amounts of information within moments. The research reviewed in this chapter supports the contention that under conditions in which scenes change far more rapidly than they ever do in normal visual scanning, and with far less predictability, our capacity to understand each brief glimpse at least momentarily is remarkably good, although our ability to remember those glimpses suffers. Not only can viewers extract the gist of a picture under these extreme conditions, they can extrapolate the scene's layout and understand the view within the context of the larger, real-world scene from which it was culled.

Table 3.2
Mean boundary score and the upper limit and lower limit of the .95 confidence interval in each serial position

	Boundary Score		Confidence Interval	
Position	M	SD	UL	LL
	All Responses			
1	–.37	.21	–.32	–.42
2	–.41	.26	–.34	–.47
3	–.38	.21	–.32	–.43
	High Confidence Responses			
1	–.44	.24	–.38	–.50
2	–.50	.29	–.43	–.57
3	–.43	.27	–.36	–.50

Boundary score range is –2 to +2.
UL, upper limit; LL, lower limit.
High confidence responses are "Sure" and "Pretty sure."
Taken from Intraub et al. 1966.

The visual system rapidly activates schematic expectations of spatial layout, given only a single glimpse of a complex scene. It also has considerable flexibility in terms of the allocation of attention, allowing viewers to minimize the effects of conceptual masking if they adopt the appropriate strategy. If these remarkable capabilities exist (a) for sequences of completely unrelated photographs (e.g., asparagus, moose, stove, man, easy chair, soldier, etc.) and, (b) under the unyielding time constraints of rapid continuous visual presentation, then it is difficult to believe that they do not take place with even greater facility during normal visual scanning. Under normal viewing conditions, a mental schema elicited by a single view could be extended and embellished as additional fixations on the scene are made. The continuity and redundancy of the successive views gleaned during visual scanning limit the likelihood of conceptual masking. The fact that both the placement and the duration of individual eye fixations can be controlled during visual scanning (something that is not possible under the artificial constraints of rapid continuous presentation) makes it seem quite feasible that a schematic representation of the spatial layout of a scene plays an important role in our comprehension of the visual environment and in the integration of briefly fixated views.

Notes

1. In the changing-ISI condition, subjects saw twice as many pictures prior to the recognition test than subjects in the repeating-ISI condition. Perhaps this, rather than online cognitive masking, resulted in greater confusion on the recognition test. This hypothesis was rejected when the ISI pictures from the changing-ISI condition were shown to subjects prior to presentation of a repeating-ISI sequence and the same results were obtained (Intraub 1984, experiment 3).

   A note about the type of equipment used in the different experiments being described here is in order. In Intraub (1984), all conditions were presented using 16 mm cine film. This differs from Intraub (1980), in which the continuous sequence was presented using cine film, and the ISI sequences were presented using a tachistoscope. The same results were obtained regardless of whether the same mode (cine film) or different modes (film and tachistoscope) were used.
2. This pattern (normalization toward the average view in the set) was predicted on the basis of the Extension-Normalization Model (Intraub 1992; Intraub et al. 1992; Intraub et al. 1996). This model provides an account of the changes in boundary memory for scenes that occur over relatively long retention intervals (e.g., 2 days).

References

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177,* 77–80.

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.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Bishop, P. J., & Intraub, H. (1996). Visual dissociation of digitized photographs. *Behavior Research Methods, Instruments and Computers, 28,* 365–371.

Carr, T. H., & Bacharach, V. R. (1976). Perceptual tuning and conscious attention: Systems of input regulation in visual information processing. *Cognition, 4,* 281–302.

Ellis, W. D. (ed. and trans). (1955). *A source book of Gestalt psychology.* London: Routledge & Kegan Paul.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307–317.

Finke, R. A. (1985). Theories relating mental imagery to perception. *Psychological Bulletin, 98,* 236–259.

Hochberg, J. (1978). *Perception* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. J. Boff, L. Kaufman, & J.P. Thomas (eds.), *Handbook of perception and human performance* (vol. 1, pp. 22:1–22:64). New York: John Wiley & Sons.

Intraub, H. (1979). The role of implicit naming in pictorial encoding. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 78–87.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 1–12.

Intraub, H. (1981a). Identification and processing of briefly glimpsed visual scenes. In D. F. Fisher, R. A. Monty, & J. W. Senders (eds.), *Eye movements: Cognition and visual perception.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Intraub, H. (1981b). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 604–610.

Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 115–125.

Intraub, H. (1985a). Conceptual masking of briefly glimpsed photographs. In R. Groner, G. W. McConkie, & C. Menz (eds.), *Eye movements and psychological processes.* North-Holland.

Intraub, H. (1985b). Visual dissociation: An illusory conjunction of pictures and forms. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 431–442.

Intraub, H. (1989). Illusory conjunctions of forms, objects and scenes during rapid serial visual search. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 98–109.

Intraub, H. (1992). Contextual factors in scene perception. In E. Chekaluk, & K. R. Llewellyn (eds.), *The role of eye movements in perceptual processes* (pp. 45–72). Amsterdam: Elsevier Science Publishers.

Intraub, H. (1977). The representation of visual scenes, *Trends in Cognitive Sciences, 1,* 217–222.

Intraub, H., Bender, R. S., & Mangers, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 180–191.

Intraub, H., & Berkowits, D. (1996). Beyond the edges of a picture. *American Journal of Psychology, 109,* 581–598.

Intraub H., & Bodamer, J. A. (1993). Boundary extension: Fundamental aspect of pictorial representation or encoding artifact? *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 1387–1397.

Intraub, H., Gottesman, C. V., & Bills, A. J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 186–201.

Intraub, H., Gottesman, C. V., Willey, E. V., & Zuk, I. J. (1996). Boundary extension for briefly glimpsed pictures: Do common perceptual processes result in unexpected memory distortions? *Journal of Memory and Language, 35,* 118–134.

Intraub, H., & Nicklos, S. (1985). Levels of processing and picture memory: The physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 284–298.

Intraub, H., and Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 179–187.

Irwin, D. E.(1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23,* 420–456.

Irwin, D. E., Brown, J. S., & Sun, J. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology: General, 117,* 276–287.

Legault, E., & Standing, L. (1992). Memory for size of drawings and of photographs. *Perceptual and Motor Skills, 75,* 121.

Loftus, G. R. (1985). Picture perception: Effects of luminance on available information and information extraction rate. *Journal of Experimental Psychology: General, 114,* 342–356.

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 435–441.

Loftus, G. R., Hanna, A. M., and Lester, L. (1988). Conceptual masking: How one picture captures attention from another picture. *Cognitive Psychology, 20,* 237–282.

s, G. R., Johnson, C. A. and Shimamura, A. P. (1985). How much is an icon worth
rnal f of *Experimental Psychology: Human Perception and Performance, 11,* 1–13.
onkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewin
mplex pictures. *Journal of Experimental Psychology: Human Perception and Performanc*
, 563–581.
onkie, G., & Zola, D. (1979). Is visual information integrated across successiv
ations in reading? *Perception and Psychophysics, 25,* 221–224.
ser, U. (1976). *Cognition and reality.* San Francisco: Freeman.
rom, L., & McClelland, J. (1992). Trace synthesis in cued recall. *Journal of Memor*
d Language, 31, 591–614.
tröm, M. (1993). Is picture memory wide-angle? *Psychological Research Bulletin,* 3
-16.
io, A. (1971) *Imagery and Verbal Processes.* New York: Holt, Rinehart, and Winsto
io, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal*
xperimental Psychology, 80, 279–285.
er, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experiment*
sychology: Human Learning and Memory, 2, 509–522.
er, M. C. (1993). Very short-term conceptual memory. *Memory Cognition, 21,* 156–16
er, M. C., & Levy, E. I. (1969). Recognition memory for pictures. *Journal of Experiment*
sychology, 81, 10–15.
ner, K., & Pollatsek, A. (1992). Eye movements and perception. *Canadian Journal*
sychology, 46, 342–376.
sberg, D. (1996). The nonambiguity of mental images. In C. Cornoldi, R.H. Logie,
. Brandimonte, G. Kaufmann, & D. Reisberg *Stretching the imagination* (pp. 119–17
lew York: Oxford University Press.
ink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need
ttention to perceive changes in scenes. *Psychological Science, 8,* 368–373.
ons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychologi*
cience, 7, 301_305.
ons, D. J. & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences,*