In D. F. Fisher, R. A. Monty and J. W. Senders (Eds.), Eye movements: Cognition and visual perception. Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1981.

III.4 Identification and Processing of Briefly Glimpsed Visual Scenes

Helene Intraub
Bucknell University

Remembering unrelated visual scenes that are presented for only a few seconds each is an easy task for the average observer. Hundreds and even thousands of pictures can be remembered with better than 90% accuracy when traditional recognition memory procedures are used (Nickerson, 1965; Shepard, 1967; Standing, Conezio, & Haber, 1970). If, however, the rate of presentation is increased to correspond to an average fixation frequency of 3 pictures/sec and faster (up to 8 pictures/sec), recognition memory suffers dramatically, approaching the level of chance (Potter & Levy, 1969). This observation that the ability to process pictorial information is severely limited when pictures are presented at the same rate that scenes are usually fixated is intriguing in that it provides an opportunity to study the nature of processes occurring at the level of a single fixation in normal viewing.

In this chapter I shall describe some recent work that has used brief pictorial presentations to simulate single fixations and sequences of briefly presented pictures to mimic the normal scanning rate. This research is concerned with two aspects of processing—identification and encoding. The first set of experiments is concerned with measuring the observer's ability to identify each briefly glimpsed scene in a sequence. The second set is concerned with the nature of encoding processes used to store briefly glimpsed pictures. This will include a discussion of the role of eye movements in encoding. The third will address the role of voluntary attentional strategies in pictorial encoding.

IDENTIFICATION OF SUCCESSIVELY PRESENTED SCENES

One possible explanation of the poor picture memory following rapid rates of presentation is that the observer is not able to identify and understand most of the pictures that are presented. Although there is considerable evidence that the gist of an isolated briefly presented visual scene is rapidly grasped by the observer (e.g., Biederman, 1972; Biederman, Glass, & Stacy, 1973; Biederman, Rabinowitz, Glass, & Stacy, 1974), it does not necessarily follow that pictures shown in a rapid continuous sequence can be so readily understood. In these sequences, the presentation rate may be similar to the normal scanning rate, but the continuity and expectancy that are characteristic of normal viewing are eliminated through the use of unrelated scenes.

To determine if identification of individual scenes is the limiting factor for memory under these conditions, Potter (1975, 1976) compared detection of a cued picture in a sequence with recognition memory for pictures presented at the same rate. Sequences of 16 magazine photographs were presented at rates ranging from 113 to 333 msec/picture. Cueing was accomplished either by showing the target picture in advance or by describing it using a brief verbal title (e.g., "a road with cars"). The rationale was that to make the match with the verbal cue, the target would have to be analyzed at a conceptual level. The proportion of targets detected was, therefore, interpreted as reflecting the minimal proportion of pictures identified during presentation. Detection was measured as a key press response falling between 250 and 900 msec following target onset. A control group was shown the same sequences, each followed by a serial recognition test that included all the pictures from the sequence and an equal number of distractors.

The results showed a marked superiority of detection accuracy over recognition memory at all rates. On the basis of only the verbal description, at a rate of 333 msec/picture, approximately 80% of all targets were detected whereas the recognition memory subjects only remembered 40% of the pictures. At the rate of 113 msec/picture, much faster than the average fixation frequency, detection accuracy was extremely good, with more than 60% of the targets detected. Only 11% of the pictures were recognized by the memory group in this condition. This was interpreted as showing that despite the lack of visual continuity during rapid presentation, pictures are momentarily understood. Following identification many are then immediately forgotten.

An alternate explanation of these results, however, holds that detection superiority may actually have been the result of expectancy (Carr & Bacharach, 1976; Neisser, 1976). Even the verbal cue may have raised probabilistic anticipations about the visual attributes of the target, thereby facilitating perception. Rather than reflecting the number of pictures momentarily identified during presentation, detection superiority may simply reflect the fact that the cued

picture was perceived more frequently than other pictures in the sequence. One way to test this possibility and determine the extent to which observers can identify unrelated pictures is to cue a picture in a sequence without providing any specific information about its visual characteristics or specific object identity. If pictures can be detected on the basis of such non-specific cues, this would indicate a striking ability of the observer to identify rapidly presented successive scenes even without the benefit of expectancy. One of the detection tasks used by Intraub (1979a) was intended for this purpose. It is described as a "negative detection" task, in which subjects are provided with a general category name prior to viewing a sequence and are instructed to detect and to describe the picture that does *not* belong to that category.

"NEGATIVE DETECTION" OF RAPIDLY PRESENTED SCENES

Sequences containing a diverse set of 11 magazine photographs from a single general category, and 1 picture (the target) that was not a member of that category were presented at rates of 114, 172, and 258 msec/picture. General categories included: transportation, house furnishings and decorations, mechanical devices, food, body parts, people, animals, fruits and vegetables, and household appliances and utensils. Pictures were selected that were as visually dissimilar as possible. For example, pictures of "animals" included creatures as diverse as a frog, a dog, a giraffe, and a butterfly. The target picture did not differ distinctively in size or in overall coloration from the other pictures in the sequence. Prior to the start of each sequence, subjects were provided with the name of the general category and were instructed to find the picture that did *not* belong (e.g., "the picture that is *not* of an animal"). In this way, the target picture was cued without providing any specific information about its visual or conceptual characteristics. The subject responded by pressing a key upon detection (reaction times were recorded) and then was required to *describe* the target picture. By requiring a description of the target, the problem of screening out false detection responses was eliminated.

Once again, the relationship between identification and memory was studied by comparing detection accuracy with a control group's recognition memory. To allow a more direct comparison than the one made in Potter's experiments, detection accuracy was compared with recognition memory for the target itself, rather than being compared with overall recognition memory for all of the pictures in the sequence. The recognition test was made more sensitive to memory for the target by introducing the following two provisions. (1) Unlike Potter's recognition test in which *all* the pictures from a sequence were tested, in the present experiment only the target picture and one other picture from the sequence were tested. This eliminated the interference that a series of relatively long tests might provide. (2) The two distractors used in the brief 4-item test were neither visually nor conceptually similar to the target. Subjects in the recognition memory condition were instructed to pay attention to each picture as it appeared and to remember as many as possible. Subjects received the recognition test immediately following each sequence. No mention of categories or "odd" pictures was made. To determine if subjects had nonetheless spontaneously categorized the pictures they were asked at the end of the experiment to write a description of the sequences they had just seen.

The results showed that subjects could detect and describe targets specified by a negative cue at all three rates of presentation. At the rate of 258 msec/picture (a rate slightly faster than the average fixation frequency), 79% of all targets were detected. When the rate was increased to 172 and 114 msec per picture, 58% and 35% of the targets were detected, respectively. Memory for the same target picture, however, did not reach the level of detection accuracy. It ranged from 19% correct in the fastest condition to 49% and 58% correct in the slower conditions, respectively. Overall, more pictures were detected than were remembered, $F(1, 84) = 9.00$, $p < .001$. At the end of the experiment, all recognition memory subjects reported that the sequences seemed to contain pictures from a general category. In fact, 87% of those subjects specifically reported noting a "category plus odd picture" arrangement. Apparently subjects had seen enough to *spontaneously* categorize the pictures during presentation. In spite of this, immediate recognition memory for the targets was inferior to detection accuracy.

These results support the hypothesis that at rates of presentation that mimic the average fixation frequency of the eye, while retention of visual scenes may be poor, the ability to momentarily identify each glimpsed scene is remarkably good. The results show that expectancy alone cannot account for the superiority of detection ability over recognition memory that was reported by Potter. Using a conservative detection task (requiring description) and a highly sensitive immediate recognition test, more targets were "negatively detected" than were remembered. Of course the results do not imply that expectancy is not important in visual perception. What they do show is that under extremely adverse conditions, in which the continuity characteristic of vision is eliminated, conceptual information specific enough to allow relatively difficult decisions to be made is available at a very early stage of processing. Apparently, the poor recognition memory performance obtained under these conditions cannot be attributed solely to the observer's inability to identify the pictures. Instead, this poor performance reflects a limitation on encoding processes necessary for retention.

ENCODING BRIEFLY GLIMPSED PICTURES

Although the gist of a briefly glimpsed picture in a sequence is apparently available at an early stage of processing, recognition memory is poor immediately following presentation. One possible explanation is that for most pictures more than one fixation is necessary for storage of a memory representation detailed enough for the picture to be recognized later. It has been suggested that the eye fixation is a special "unit of encoding" (Loftus, 1972). According to this view, encoding takes place during the fixation, while the picture is physically present—perhaps specifically limited to the early part of the fixation (Loftus, 1976). The first fixation is seen as providing the observer only with general information about the picture; each additional fixation is thought to increase the probability that an informative detail will be encoded, thereby increasing the likelihood that a picture will be recognized later (Loftus & Kallman, 1979). The poor memory performance obtained following rapid presentation is attributed to the fact that only one simulated fixation is possible per picture.

There is a growing body of evidence, however, that disputes the notion that encoding is limited to the duration of the fixation. Rosenblood and Pulton (1975), for example, presented 400 pictures for as little as 80 msec each with a 5 sec blank interstimulus interval (ISI). Subjects recognized 74% of the pictures. This performance is far superior to that obtained when briefly presented pictures are shown in a continuous sequence. Intraub (1980) directly compared recognition memory for 150 magazine photographs presented for 6 sec each with no ISI, and the same pictures presented for only 110 msec each followed by a 5890 msec blank ISI. This drastic reduction in stimulus duration resulted in a surprisingly small decrease in recognition memory from 94% to 77% correct, even though the number of fixations probably dropped from about 18 to 1. When the same pictures were presented in a continuous sequence at a rate of 110 msec/picture, recognition memory dropped dramatically, with only 21% recognized. In another experiment, pictures were presented for 110 msec each, followed by blank ISIs of 1390, 620, 385, 165, or 0 (no ISI) msec. Recognition memory decreased from 92% to 83%, 74%, 57%, and 19% correct, respectively (Intraub, 1979b). This indicates that encoding is not limited to the duration of a fixation, although the nature of the encoding process is not readily apparent.

The decrease in memory obtained when the ISI between briefly presented pictures is reduced can be explained equally well by two general encoding hypotheses. (1) Encoding is an all or nothing phenomenon. For a given observer, a particular picture requires a fixed amount of time to be encoded in memory. If enough time is not allowed, the picture will be forgotten. As the time between pictures is diminished, although many pictures can be momentarily identified, fewer pictures can be encoded. (2) Encoding is a continuous process. Following identification, an increasing number of pictorial details will be stored. As the time between pictures is diminished, fewer pictures will be stored in enough detail to pass a recognition threshold at the time of the test. While both hypotheses predict a decrease in recognition memory when presentation rate is increased, they differ in their predictions regarding the stored representations of the pictures. According to the all or nothing interpretation, given equal stimulus duration, those pictures that are recognized following a rapid rate of presentation should be remembered in as much detail as pictures that are recognized following a slower rate of presentation. On the other hand, the continuous encoding hypothesis predicts that as the time between pictures is reduced, not only will fewer pictures be remembered but they will be remembered in less detail. A traditional recognition test that uses dissimilar distractors cannot distinguish between these two hypotheses because a minimal amount of information might be sufficient to elicit a recognition response. To avoid this problem, the following experiment was conducted.

RETENTION OF DETAIL FOLLOWING BRIEF PICTORIAL EXPOSURES

To determine if less is remembered about each picture as the time between pictures is diminished, Intraub (1980) used a recognition test with two levels of difficulty. Subjects first were required to indicate whether or not they recognized a picture. Following that decision they then had to determine if the picture was in the same orientation as in the inspection sequence or if it was mirror reversed. According to the all or nothing hypothesis, with a constant stimulus duration if a picture is recognized, the ability to detect a reversal should remain the same regardless of the ISI. According to the continuous encoding hypothesis, the ability to determine that a recognized picture is reversed should decrease as the time between pictures is diminished.

Twenty magazine photographs were presented for 110 msec each with blank ISIs of 4890, 1390, 620, 385, or 0 (no ISI) msec, or for 5 sec each with no ISI. Ten subjects participated in each condition. They were instructed to pay attention to each picture as it appeared and to try to remember as many as possible. *No mention of mirror reversal was made at this time.*

Following presentation of the sequence, a serial recognition test was administered that contained 16 pictures from the inspection sequence (the initial and final pairs of pictures were not tested) and 16 dissimilar distractors (new pictures). Half of the 16 target pictures were mirror reversed for half the subjects in each condition. Subjects were informed that some of the pictures in the test would be mirror reversals of pictures they had seen in the inspection sequence. They were instructed to respond "yes" if they recognized a picture regardless of the picture's orientation, and "no" if not. Following a "yes" response they were told to respond "reversed" if they thought that the picture was mirror reversed and "normal" otherwise.

The proportion of pictures recognized ("yes" responses, corrected for guessing) dropped only from .96 to .84 when stimulus duration was reduced from 5 sec to 110 msec with a 4890 msec ISI. Recognition memory remained the same in the 1390 msec-ISI condition (.84), but decreased to .61, .48, and .20 as the ISI was further reduced (see Fig. 1). At all rates, however, recognition was better than chance. Reversing a picture did not affect the subject's ability to recognize it

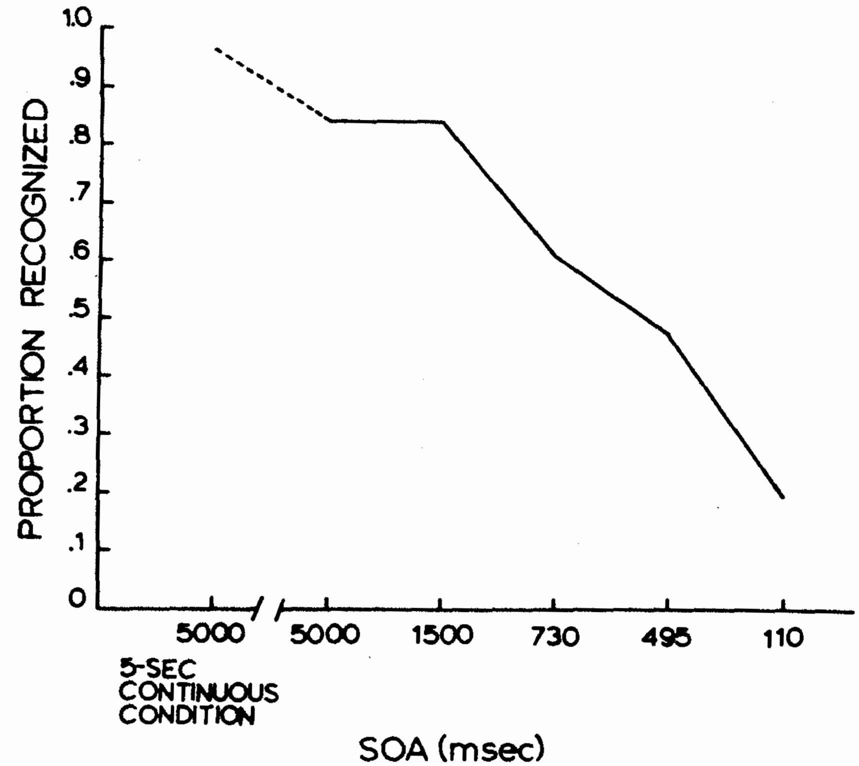

FIG. 1. The proportion of 110 msec pictures recognized as a function of stimulus onset asynchrony (SOA) and the proportion of pictures recognized in the 5 sec/picture control condition.

at any rate (an observation also reported by Standing, Conezio, & Haber, 1970, when pictures were presented for a few seconds each).

When a picture was recognized, subjects in the 5 sec continuous condition were able to detect a reversal 81% of the time. Reducing stimulus duration to 110 msec with a 4890 or 1390 msec ISI lowered but did not eliminate this ability (see Fig. 2). As the time between pictures diminished, the ability to detect reversal decreased sharply and was not significantly better than chance at the three fastest rates.

These results demonstrate that rather than being an all or nothing process confined to the initial part of a fixation, pictorial encoding is a continuous process. Although subjects could successfully recognize some pictures at each rate of presentation, the ability to determine that they were reversed was eliminated as the time between pictures was reduced, thus indicating reliance on a less detailed memory representation. This reduction in reversal detection occurred even though stimulus duration was held constant at 110 msec preventing the

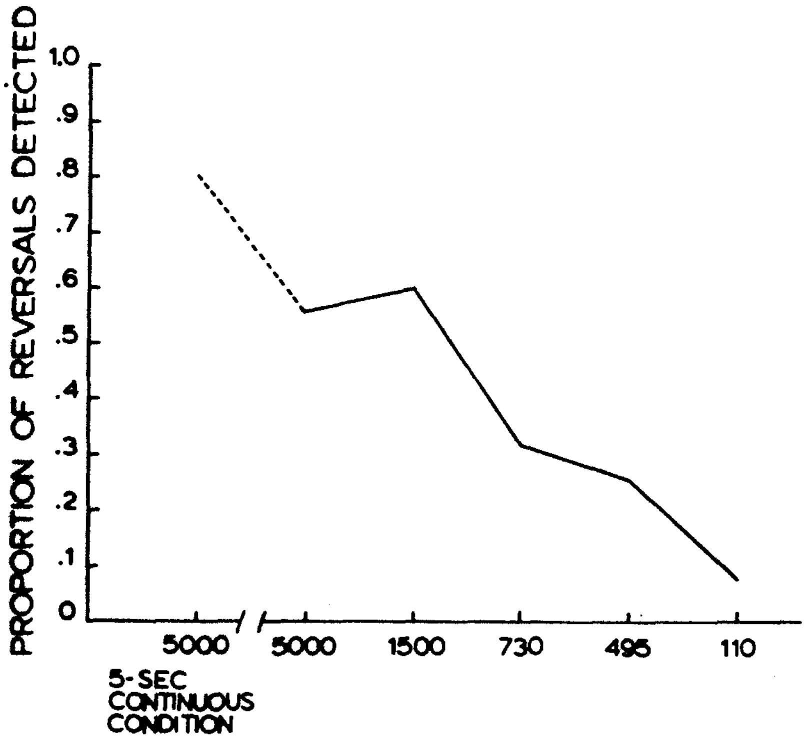

FIG. 2. The proportion of reversals detected as a function of stimulus onset asynchrony (SOA) with a 110 msec stimulus duration and the proportion detected in the 5 sec/picture control condition.

subject from making any additional eye fixations on the pictures during encoding. Since the pictures contained no alphanumeric characters, it is unlikely that left-right orientation of a picture was relevant to its meaning. Even so, after a brief glimpse, subjects were not merely limited to remembering the "gist" of a picture but frequently remembered the objects and scenes in enough detail to determine that they were mirror reversed provided that the time between pictures was long enough. To their own surprise, subjects reported that certain pictures simply *looked* backwards.

As would be expected, memory was somewhat better when pictures were presented for a full 5 sec each than when they were presented for only 110 msec followed by a blank 4890 msec ISI. In the former case the subject could continually scan the picture while encoding was taking place, perhaps storing additional details (e.g., Loftus & Kallman, 1979). The present results show that encoding of visual detail is not confined to the duration of the stimulus, nor is it necessarily dependent on the number of eye fixations made on a picture. Encoding of

information necessary for reversal detection continued beyond the period of iconic persistence, and in fact beyond a 620 msec ISI. Under these conditions of presentation it appears that encoding of each picture is terminated when processing of the next picture begins. Does this mean that in normal vision, encoding of each "fixated" scene is terminated by the onset of the next fixation? Some recent work concerned with the role of attentional processes in encoding briefly glimpsed pictures suggests that this is not necessarily the case.

ATTENTIONAL STRATEGIES IN ENCODING BRIEFLY GLIMPSED PICTURES

Potter and Levy (1969) suggested that processing of each picture in a sequence is terminated by the appearance of the "next substantial visual event." Recent research, however, suggests that to some extent there may be voluntary control over whether processing will continue. Potter (1976) demonstrated that when briefly exposed pictures were interspersed with presentation of a colorful visual noise mask, recognition memory far surpassed that obtained when the same pictures were presented in a continuous sequence. Intraub (1980) presented pictures for 110 msec each followed by an ISI that either contained a blank field or a familiar picture that repeated throughout the sequence. Subjects were instructed to attend to the briefly presented pictures. Presentation of a repeating picture during the ISI interfered only minimally with recognition memory and did not affect the ability to detect reversal. Evidently, processing of a picture with a duration of at least 110 msec can continue despite the onset of a meaningless visual noise mask or a repeating picture.

In both cases, however, the ISI contained a familiar, expected visual event. During continuous presentation, each picture is followed by a new picture. To determine if observers can effectively ignore the onset of a new meaningful visual event, as they seem to be able to do with a familiar one, Intraub (1979c) presented pictures for 110 msec each with a 1.5 sec ISI that contained a blank field, a repeating picture, or a new picture each time. Again, subjects were instructed to attend to the briefly presented pictures. Recognition accuracy for the blank and repeating picture conditions was 89% and 80%, respectively. When a *new picture* was presented during the ISI each time, even though recognition memory for the briefly glimpsed pictures decreased significantly (dropping to 63% correct), it did not approach the low level of performance obtained following rapid continuous presentation. This level was reached when the subject was instructed to attend to the "long" pictures (the ISI pictures). As attention instructions were changed to place emphasis on "brief" pictures, all pictures, or "long" pictures, recognition memory for the brief pictures decreased dramatically from 63% to 12% correct. At the same time, memory for the long pictures (that had been presented for 1.5 seconds each) increased from 54% to 87%

correct. This indicates that not only can encoding continue in spite of the occurrence of a new visual event, but that to a large degree, the allocation of attention to sequentially presented visual information can be controlled voluntarily. In the case of rapid sequential presentation of unrelated pictures (no ISI), encoding of each picture may be disrupted when processing is initiated on the next pictures in the sequence (Intraub, 1980; Potter, 1976).

SUMMARY AND IMPLICATIONS

Without expectancy and continuity that characterize visual scanning, observers were able to conceptually identify unrelated, successively presented pictures at rates that mimic average fixation frequency. Identification did not necessarily result in retention. Detection accuracy (even with the use of a "negative" cue) was superior to immediate recognition memory for the same pictures. The ability to *momentarily* identify glimpsed scenes may function as a monitor in vision. For example, the ability to rapidly identify each fixated "scene" may play a role in controlling placement of subsequent fixations.

Apparently, following identification of a scene, the quality of the memory representation depends in large part on how soon encoding is disrupted. As the time between briefly glimpsed pictures was increased, the ability to remember a picture's left-right orientation increased. This superior memory was obtained without additional fixations having to be made on the picture. Although the factors that determine when encoding will be disrupted are not yet fully defined, it is clear that to a large degree the observer can voluntarily control the encoding process. When briefly glimpsed pictures are presented with an ISI that contains a redundant visual event, they are remembered nearly as well as when the ISI contains a blank field. Under conditions in which the ISI contains a new, unrelated picture each time, encoding of briefly presented pictures can extend beyond a single fixation if the observer is instructed to remember the briefly glimpsed pictures and ignore the intervening ISI-pictures. It seems likely that when scanning a coherent visual scene in which successive glimpses are related, encoding will not be limited to the duration of a fixation. The results suggest that the observer can rapidly assess the importance of each fixated "scene" during normal viewing and adjust the extent of encoding to be carried out. Perhaps these overlapping encoding processes play a role in the integration of successive fixations.

ACKNOWLEDGMENT

This research was supported in part by Advanced Research Projects Agency Contract MDA 903-76-C-0441 to Massachusetts Institute of Technology.