

Tachistoscopic Simulations of Eye Fixations on Pictures

Geoffrey R. Loftus
University of Washington

In three picture recognition experiments, complex pictures were presented during a study phase, each presentation consisting of n sequential masked presentations and each presentation lasting d msec. This procedure was designed to mimic a series of eye fixations over a picture, but with number and duration of fixations under experimental control. With n held constant, subsequent recognition memory performance increased with increasing d up to 400 msec. With d held constant, performance increased with increasing n only to the degree that an additional presentation of a picture was used to fixate a novel portion of the picture. These results, and those of Loftus's 1972 experiment, suggest a model of picture encoding that incorporates the following propositions: (a) A normal fixation on a picture is designed to encode some feature of the picture. (b) The duration of a fixation is determined by the amount of time required to carry out the intended feature encoding. (c) The more features are encoded from a picture, the better the recognition memory will be from the picture. Additionally, the results of the present experiments imply that the events that constitute encoding within a fixation proceed in a fixed, relatively immutable order.

When a human being scans a static visual scene, the eye movement pattern is broken up into a series of periods during which the eye is relatively stationary (fixations) separated by quick jumps of the eye from place to place (saccades). Acquisition and processing of visual information occurs almost exclusively during the fixation periods, which are typically about 300 msec in duration. Accordingly, a large amount of research has been concerned with investigating properties of the activity that occurs within a fixation. This article is limited to a discussion of the information acquisition that occurs during fixations of complex pictures.

Eye Fixations and Recognition Memory for Pictures: The Asymptote

In 1972, I reported a series of picture recognition experiments whose results allowed

This research was supported by National Science Foundation Grants BNS76-23367 and BNS79-06522 to Geoffrey Loftus. I thank Charles Traneer, Mary Kolpacoff, and Steven Yantis for conducting Experiments 1, 2, and 3, respectively. Gail Goodman, Eileen Kowler, and Ellen Markman provided helpful comments on an earlier draft of the manuscript.

Requests for reprints should be sent to Geoffrey R. Loftus, Department of Psychology, University of Washington, Seattle, Washington 98195.

inferences about the time course of information acquisition within an actual eye fixation. In the first experiment of the Loftus (1972) series, target pictures were presented in pairs during a study phase. Each pair was displayed for 3 sec, and the subject was free to look back and forth over the two members of the pair. Prior to presentation, each member of the pair was assigned a value ranging from 1 to 9 points. A picture's point value eventually translated into money gained if the picture was correctly recognized at the time of test or money lost if the picture was not correctly recognized. It turned out that the number of fixations made on a target picture at the time of study was highly predictive of the probability that the picture would eventually be recognized. It also turned out that with number of fixations held constant, average fixation duration was longer for higher valued than for lower valued pictures. However, recognition memory was not affected by the value manipulation, which means that the additional time per fixation spent on the higher valued pictures was not being used to acquire information that would be useful in the eventual recognition test.

In the second experiment of the Loftus (1972) series, pictures were viewed singly

during a study phase, and exposure duration was varied from 300 to 5,000 msec. Two results of interest emerged. First, for a fixed exposure time (e.g., 3,000 msec) recognition performance increased with increasing numbers of fixations; that is, many short fixations led to better performance than did fewer longer fixations. Second, with number of fixations held constant, performance did not vary with exposure time. For example, 12 fixations made during a 3,000-msec exposure resulted in the same recognition performance as did 12 fixations made during a 5,000-msec exposure.

These results suggest that within a single eye fixation, all the information germane to subsequent recognition memory is acquired relatively early in the fixation. That is, total information acquired seems to asymptote after a time that is shorter than the duration of the fixation. Two possible explanations of this state of affairs were offered. The first incorporates the notion that the onset of new retinal stimulation initiates a "visual information processing" routine that operates for only a fixed period of time (e.g., 100 msec) and then quits until another change in retinal stimulation takes place. The second explanation incorporates a form of item-selection hypothesis which is that a fixation is designed to encode some particular feature of the fixated area, and that features vary in the amounts of time required to encode them. Thus, a picture with quickly encodable features would allow the encoding of more features in a given time than would pictures with features requiring longer per-feature encoding times. Assuming that (a) more encoded features result in better memory performance and (b) any fixation lasts at least the amount of time necessary to encode the feature at hand, the results of the Loftus (1972) experiments would readily follow.

Note that these two explanations differ with respect to their implications about causal relations. According to the first explanation, a particular level of memory performance is viewed as caused by a particular number of fixations at the time of the original viewing; that is, a particular number of fixations is both necessary and sufficient for a particular level of performance. According to the second explanation, both number of

fixations and memory performance are caused by the physical configuration of the picture. Thus, a particular number of fixations is necessary but not sufficient for a particular level of performance.

Since fixation durations were not under experimental control in the Loftus (1972) experiments, these two explanations could not be distinguished. In the experiments reported here, a picture recognition procedure is introduced that is designed to experimentally manipulate both fixation duration and number of fixations on a picture. In the study phase of this "multiple-flash paradigm," a picture was presented (flashed) n sequential times at durations of d msec per flash. Each of the n presentations was followed by a pattern mask designed to eliminate the iconic image of the picture. In this procedure, n , the simulated number of fixations, and d , the simulated fixation durations, were both experimenter-controlled variables. If the "self-stopping program" explanation is correct, then increasing flash duration (d) beyond some critical value should not lead to increases in memory performance. If, in contrast, the item-selection explanation is correct, then any increase in flash duration should lead to a corresponding increase in eventual performance, since increasing flash duration should increase the probability that any given feature can be encoded.

All of the present experiments were yes/no recognition experiments in which the major independent variable was study condition, as defined by combinations of n and d . Conceptually, the three experiments were quite similar, the major differences being that (a) different study conditions (n - d combinations) were used in the three experiments, (b) the randomization procedures differed slightly, and (c) eye movements were recorded during the study phase of Experiment 3.

Experiment 1

In Experiment 1, all pictures were displayed during a study phase for a total duration of 400 msec. However, these 400 msec were broken up in four different ways, resulting in n - d combinations of 1-400, 2-200, 4-100, and 8-50. Recall that the Loftus

(1972, Experiment 2) results indicated that increased numbers of (shorter) fixations during a given exposure time resulted in superior recognition performance. If there is a causal relationship between number of fixations and performance, then the same result ought to occur when number and duration of fixations are experimentally controlled, as in the present experiment; that is, performance ought to increase with increasing n .

Method

Subjects. Subjects were eight University of Washington undergraduate volunteers who responded to a newspaper advertisement. They were paid \$2 per hour for participating, and were tested individually.

Stimuli. Stimuli consisted of 320 35-mm slides divided into four trays of 80 slides per tray. The pictures were naturalistic color photographs of landscapes, cities, houses, sunsets, and so on. Pictures containing people or writing of any sort were not used. Pictures were back projected on a screen approximately 68 cm from the subject's eyes. They subtended a visual angle of $13.2^\circ \times 9^\circ$.

Apparatus. Pictures were displayed with a Kodak random-access projector. At study, a second, standard Kodak Carousel projector was used to display the 300-msec pattern mask, which immediately followed each flash of a target picture. The relationship of target and mask was such that when the two were displayed concurrently, no information could be acquired from the target.

Design. Each of the four slide trays constituted an independent replication in the following sense. Pictures in a given slide tray were divided into two sets corresponding to slides in odd-numbered and even-numbered slots. Either the 40 odds or the 40 evens were displayed as targets in a study session, followed immediately by a yes/no recognition test on all 80 slides presented in random order. During the study phase, each of the 40 slides fell into one of the four n - d combination study conditions listed above. The ordering of conditions of study was random, with the constraint that 10 slides fall into each of the four conditions. Each target picture was rotated through the four study conditions over four subjects; in addition, each slide served equally often as target and as distractor. This procedure required eight subjects.

Procedure. After receiving instructions about study and test procedures, a subject was shown examples of the four study conditions. After questions, the experiment proper began. A study phase involving 40 trials with a given slide tray was followed by a test of all 80 pictures in that tray, and this study-test procedure was followed for each of the four trays in sequence. During the study phase, subjects were not informed of the study condition of a slide on any given trial. When n , the number of flashes, was greater than 1, there was a 300-msec interval between the offset of the mask following flash i and the onset of the flash $i + 1$. There was an interval of approximately 2 sec (the time needed to ad-

vance the slide projector and reset the timer) between trials.

The test phase immediately followed the study phase for each of the four trays. At the beginning of each test phase, subjects were reminded that the correct response was yes for half the test trials and no for the other half. They were urged to respond yes and no about equally often. The test phase was self-paced; subjects typically took no longer than about 6 sec to respond on each trial. No feedback was provided during the test phase.

Results

The results for each of the four study conditions are shown in Table 1. A d' score was computed for each subject for each condition by subtracting the z score corresponding to the false-alarm rate from the z score corresponding to the hit rate. All data presented are means of the individual d' scores.

There is no hint of an increase in performance, as the total study duration of 400 msec is divided into increasing numbers of increasingly shorter flashes. In fact, the 1-400, 2-200, and 4-100 conditions show virtually identical performance and the 8-50 condition is substantially worse than the other three.

Experiment 2

In Experiment 1, flash duration and number of flashes were confounded, since total study time was held constant across conditions. In Experiment 2, five study conditions (n - d combinations) were chosen to produce a partial factorial combination of n and d .

Method

Subjects. Subjects were 10 University of Washington undergraduate volunteers who responded to a newspaper ad. They were paid \$2 for participating in the experimental session. None had participated in Experiment 1.

Table 1
Recognition Memory Performance (d') for the Four Study Conditions of Experiment 1

Study condition	d'
1-400	1.70
2-200	1.72
4-100	1.66
8-50	0.84

Note. 95% confidence interval is $\pm .12$.

Stimuli and apparatus. Stimuli and apparatus were identical to those used in Experiment 1.

Design. During the study phase, each of the 40 slides in a given tray fell into one of five study conditions. In terms of n - d combinations, these five conditions were 2-50, 4-50, 1-100, 2-100, and 1-200. Within a tray, the order of conditions was random, with the constraint that 8 slides fall into each of the five conditions. Each target picture was rotated through each of the five conditions over subjects; in addition, each picture served equally often as target and distractor. This required 10 subjects.

Procedure. The procedure was the same as in Experiment 1.

Results

The results, shown in Figure 1, panel A, depict recognition performance (d') as a function of d , the flash duration. The curve parameter is n , the number of flashes. It is evident that performance increases as a function of d . Although higher values of n tend to show higher performance, the effect is not very dramatic. The dotted and dashed lines form "isototal time" curves; that is, they connect data points of conditions in which total presentation time ($n \times d$) is constant. With total time held constant, performance *decreases* as a function of n , the number of flashes.

Discussion: Experiments 1 and 2

A major finding of the Loftus (1972) experiments was that with exposure time held constant, performance increased with increasing numbers of fixations. When eye fixations are simulated tachistoscopically and their durations experimentally controlled, all traces of this phenomenon disappear. Moreover, the results of the Loftus (1972) experiments as well as those of Sperling (1963), Allport (1968), and Sperling, Budiansky, Spivak, and Johnson (1971) all suggest that within a fixation, visual information processing ceases fairly early, that is, acquired information reaches asymptote soon after the start of a fixation. There was no exposure time asymptote in the data of Experiment 2.

The suggestion therefore is that given a fixed exposure time, there is no causal relationship between number of fixations on a picture made during that time and amount of information stored about the picture. This conclusion, however, must be tempered by

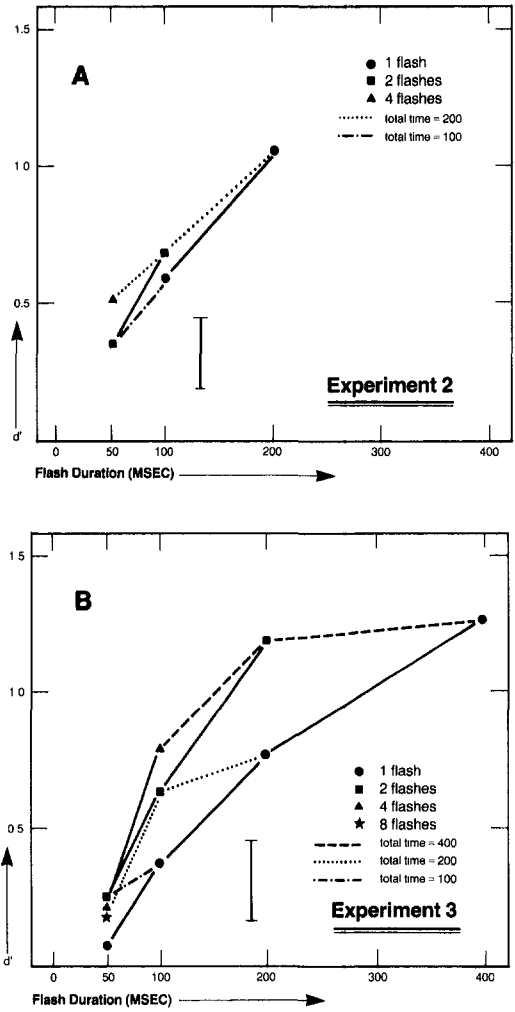


Figure 1. Panel A depicts recognition performance (d') as a function of d , the flash duration for Experiment 2. (Separate curves are drawn for different values of n , the number of flashes: circles for $n = 1$ and squares for $n = 2$. The triangle represents the [$d = 50, n = 4$] condition. Dotted and dashed lines connect data points for conditions in which total exposure time [$n \times d$] is constant.) Panel B depicts analogous data for Experiment 3. (The triangles form the curve for the $n = 4$ flash conditions. The star represents the [$d = 50, n = 8$] condition.) In both panels, the vertical bar represents the 95% confidence interval around the data points.

the question of how much the tachistoscopic exposures of the present paradigm are really simulations of eye fixations. In particular, an eye fixation has the very salient property that it shifts the gaze to a new place in the picture. There is no guarantee that this oc-

curred in the present paradigm; it is entirely possible that subjects were simply holding their eyes steady throughout all n flashes.

Experiment 3

Experiment 3 was designed to test this possibility. The paradigm of Experiments 1 and 2 was used with the additional feature that eye movements were recorded during the study phase.

Method

Subjects. Subjects were 20 University of Washington undergraduates recruited via a newspaper advertisement. All were tested individually and were paid \$2 per session for participating. None had participated in Experiments 1 or 2.

Stimuli and apparatus. The stimuli and stimulus display apparatus were identical to those used in Experiments 1 and 2. During the study phases, eye movements were recorded using a corneal-reflection-based system described in detail by Loftus (1979). This system produced as output a video recording of the scene viewed by the subject, on which was superimposed a small dot corresponding to eye position. The system is accurate to approximately $\pm .5^\circ$ in both the vertical and the horizontal.

Procedure. Ten study conditions were used. Again, with study condition specified as a particular n - d combination, the conditions were 1-50, 2-50, 4-50, 8-50, 1-100, 2-100, 4-100, 1-200, 2-200, and 1-400.

In contrast to Experiments 1 and 2, the study phase for each slide tray in Experiment 3 was blocked by condition. All 10 conditions were incorporated in each of the 4 slide trays; thus the 40 study slides for a given tray were shown in 10 blocks of 4 trials per block, 1 block corresponding to each condition. Prior to the start of a block, subjects were informed as to the study condition of trials in the upcoming block. On each trial prior to target presentation, the subject fixated a central point, and if necessary, the eye movement recorder was realigned. Such realignment typically took no more than 5 sec.

As in Experiments 1 and 2, each slide served equally often in all 10 study conditions and equally often as target and distractor. Thus, 20 subjects were required for the experiment.

Results

Performance for the 10 conditions is shown in Figure 1, panel B. Again, d' is shown as a function of flash duration, and the curve parameter is number of flashes. Again, dotted and dashed lines represent isototal time curves. The data of Experiment 2 are replicated in the sense that performance for the 100- and 200-msec conditions increases

strongly as a function of flash duration and only modestly as a function of number of flashes. (Of some interest is the fact that the 50-msec conditions show virtually *no* increase as a function of number of flashes.) An examination of the isototal time curves indicates that with total time held constant, performance decreases as a function of number of flashes.

Eye movement data. Figure 2 shows the data that result when eye fixations are taken into consideration. Again, d' was computed individually for each condition for each subject, and these results were averaged. Panel A shows data for flash durations of 100-400 msec, whereas panel B shows data for the 50-msec conditions.

Consider Figure 2, panel A. For a given number of flashes, n , the subject could have looked at a minimum of one and a maximum of n different places in the picture. (The one exception is that the 1-400 condition allowed more than one fixation.) The data shown in Figure 2, panel A, are conditionalized on both n and number of places looked. For the 4-100 condition, the column of four triangles, labeled "1 place . . . 4 places," above the 100-msec abscissa point constitutes the data for the four possible looked-at places. The 400-msec-condition data point represents instances in which only a single fixation was made during the 400-msec flash.

There are several noteworthy aspects of these data. First, the strong increase in performance across flash duration is still seen to occur. Second, performance depends quite dramatically on the number of places looked at. Even more striking is the fact that when number of places looked at is held constant, the effect of number of flashes vanishes, thereby indicating that additional flashes are only useful insofar as they permit acquisition of information from additional portions of the picture. This notion is perfectly reasonable in the sense that the structure of the retina is such that fine-detail encoding can be performed only on the small (approximately $2^\circ \times 3^\circ$) area of the picture that falls on the foveal region. As shown by Nelson and Loftus (1980), information pertinent to subsequent recognition memory seems to be acquired only from this small region during a given fixation. Hence, the more places

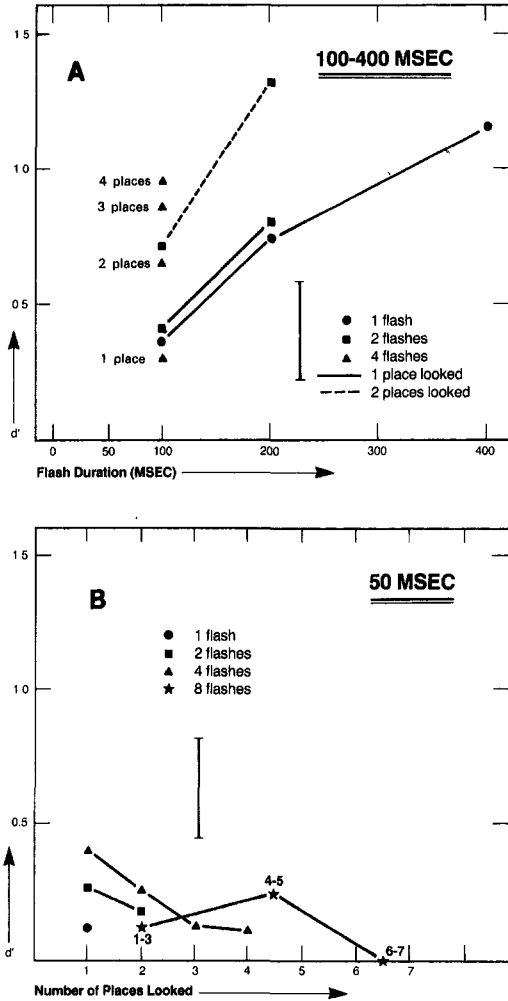


Figure 2. Panel A depicts recognition performance (d') as a function of d , the flash duration for Experiment 3. (Data from the 50-msec conditions are excluded. Separate curves are drawn for n , the number of flashes, and for the number of different places looked at in the picture across the n flashes. Circles, squares, and triangles represent n values of 1, 2, and 4, respectively. Solid and dashed lines represent one and two places looked at, respectively. Triangles represent data from the [$d = 100$, $n = 4$] condition, in which the number of places looked at ranged from one to four.) Panel B depicts data for the 50-msec conditions of Experiment 3. (Curves represent d' as a function of number of different places looked at. Circles, squares, triangles, and stars are for n values of 1, 2, 4, and 8, respectively). In both panels, the vertical bar represents the 95% confidence interval around the data points.

looked at in a picture, the more information is acquired from the picture.

Figure 1, panel B, suggests that the 50-msec flashes were somewhat odd in that

there was very little, if any, increase in performance over number of flashes. Figure 2, panel B, supports this suggestion, showing performance as a function of number of places looked at for the 50-msec flashes. Different curves are shown for the one-, two-, four-, and eight-flash conditions. Because of missing data points, the eight-flash curve was collapsed over the one to three places, four to five places, and six to eight places looked at. Inspection of these curves indicates that in contrast to their longer duration brethren, 50-msec glimpses of additional portions of the picture did not lead to increases in memory performance. If anything, the curves declined somewhat.

General Discussion

The Relationship Between Fixation Duration and Picture Recognition

The principal finding that emerges from these experiments is that when a picture is shown in a series of tachistoscopic flashes, information useful in a subsequent recognition test is, on the average, extracted continuously from each flash at least up to a flash duration of 400 msec. This finding has direct relevance for the findings of Loftus (1972) that recognition memory performance for pictures is independent of fixation duration at the time of original study. In particular, the following assumptions seem adequate to account for performance in both the Loftus (1972) and the present experiments:

1. The purpose of fixation is to encode some feature from the fixated area.
2. An ordinary fixation continues at least as long as necessary to carry out required encoding.
3. The more features encoded from a picture, the better subsequent recognition memory performance is for the picture.

Implications of Variation in Feature-Encoding Time

Under ordinary viewing conditions, a strong determinant of the number of fixations given a particular picture during a fixed time period is the physical nature of the pictures. Pictures that have easy-to-encode features will receive many short fixa-

tions, whereas pictures that have difficult-to-encode features will receive fewer but longer fixations. This, along with Assumption 3 above, accounts for the relationship between number of fixations and recognition memory performance reported by Loftus (1972). Since more encoded features lead to better memory performance, ease of feature encoding should also affect picture memory. For example, encoding of features from pictures that are upside down or out of focus would likely take longer than encoding of features from normally presented pictures. As shown by Dallett, Wilcox, and D'Andrea (1968), pictures presented in these unusual ways do indeed show poorer recognition performance.

Experimental Control of Fixation Duration

When fixation duration is artificially controlled, as in the present paradigm, a different story emerges. In particular, any controlled-duration fixation has some probability of being terminated prior to completion of the intended encoding during that fixation. The shorter the fixation, the greater the probability this will occur. Therefore, shorter controlled fixations lead to fewer encoded features on the average and poorer memory performance.

Automaticity of Encoding Within a Fixation

The eye movement data in Figure 2, panel B, indicate an additional constraint on the relationship between eye fixations on pictures and subsequent recognition memory: A fixation is useful only to the degree that it falls on a novel place in the picture. Given earlier arguments that encoding is sometimes cut off prematurely during a tachistoscopic exposure, this notion yields an interesting implication. One might suppose that an encoding process prematurely terminated on flash i could then be completed on flash $i + 1$. Evidently this is not the case, because if it were, then average performance would increase with increasing numbers of flashes whether or not the additional flashes were used to look at new places in the picture. Rather, it appears that when a new

place is not fixated on flash $i + 1$, the same information acquired on flash i is reacquired on flash $i + 1$. Apparently, the sequence of events that constitutes encoding within a tachistoscopic flash is relatively automatic. That is, the sequence seems to start from the beginning following the onset of new retinal stimulation and proceed in an immutable order.

Processing During 50-msec Flashes

A sequence of 50-msec flashes leads to qualitatively different performance than does a sequence of 100- or 200-msec flashes. This assertion receives dramatic support from the data in Figure 2, panel B, which indicate no improvement in performance with increasing numbers of flashes or places looked at. This is not merely a floor effect, as overall 50-msec performance exceeds chance.

There are at least two reasonable explanations for this finding. The first is that 50 msec is not sufficient time to carry out the peripheral scanning necessary to determine where a subsequent fixation in the picture should be. Informal analysis of the videotape records provides some confirmation for this notion. In general, an eye movement during a sequence of 100-msec or 200-msec flashes brought the gaze to some particular, predictable point in the picture, such as a prominent object. However, eye movements during a sequence of 50-msec flashes did not share this property; rather, they wound up on relatively random places in the picture. Thus, it would seem that a fixation on a new place in the picture is not sufficient for the acquisition of useful information and that another fixation on a place that is in some sense "informative" is necessary.

The second explanation revolves around acquisition of "gist information." Biederman and his colleagues (Biederman, 1972; Biederman, Rabinowitz, Glass, and Stacy, 1974) demonstrated that the gist of a picture can typically be acquired within about 100 msec following the picture's onset. The present 50-msec results would obtain if (a) 50 msec were not a sufficiently long exposure for the subject to acquire the gist of the picture and (b) gist information were required in order for subsequent processing to proceed.

Neither of these possibilities provides a

satisfactory explanation for why the 50-msec performance is, overall, above chance. At present, there are only speculative explanations for this, which will not be discussed here.

Masking Considerations

An implicit assumption in this article is that the effect of a pattern mask is simply to halt target processing, as was suggested by Sperling (1963). However, a plausible alternative is that an iconic image of the picture persists into the masking period; that is, that masking does not entirely halt picture processing (cf. Eriksen, 1980). This possibility cannot be entirely ruled out. However, the following considerations weigh against its importance:

1. The sort of pattern mask used in the present experiments is precisely the kind that would likely produce interruption-type masking (Eriksen, 1980).

2. As noted, when target and mask were superimposed, no information could be extracted from the target. Thus, the paradigm passed the most basic test suggested by Eriksen (1980).

3. In past experiments we have gathered picture memory data using the same mask as that used in the present experiments. These data have been well fit by quantitative models that assume that the onset of the mask terminates picture processing.

4. Even if masking did *not* terminate processing, the present conclusions would still hold, since nontermination of processing would improve performance with increasing values of n .

Poststimulus Processing

An as yet unresolved issue in the picture memory literature is the degree to which processing of a picture continues following the disappearance of the physical stimulus

(cf. Intraub, 1980; Shaffer & Shiffrin, 1972). It is possible, although by no means certain, that in the present experiments processing continued on the memorial representation of the picture during the intervals between flashes. It should be noted, however, that the same logic suggested in Points 3 and 4 in the Masking Considerations section applies to this issue as well. Most importantly, post-stimulus processing, to the degree that it occurs, would work against the conclusion that increasing numbers of shorter flashes do not boost memory performance.

References

- Allport, D. A. The rate of assimilation of visual information. *Psychonomic Science*, 1968, 12, 231-232.
- Biederman, I. Perceiving real-world scenes. *Science*, 1972, 177, 77-80.
- Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. On the information extracted from the scene at a glance. *Journal of Experimental Psychology*, 1974, 103, 597-600.
- Dallett, K., Wilcox, S. G., & D'Andrea, L. Picture memory experiments. *Journal of Experimental Psychology*, 1968, 76, 312-320.
- Eriksen, D. W. The use of a visual mask may seriously confound your experiment. *Perception & Psychophysics*, 1980, 28, 89-92.
- Intraub, H. Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, 6, 1-12.
- Loftus, G. R. Eye fixations and recognition memory for pictures. *Cognitive Psychology*, 1972, 3, 525-551.
- Loftus, G. R. On-line eye-movement recorders: The good, the bad, and the ugly. *Behavior Research Methods and Instrumentation*, 1979, 11, 188-191.
- Nelson, W. W. & Loftus, G. R. The functional field of view during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, 6, 391-399.
- Shaffer, W. O., & Shiffrin, R. M. Rehearsal and storage of visual information. *Journal of Experimental Psychology*, 1972, 92, 292-295.
- Sperling, G. A model for visual memory tasks. *Human Factors*, 1963, 5, 19-31.
- Sperling, G., Budiansky, J., Spivak, J., & Johnson, M. C. Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science*, 1971, 174, 307-311.

Received October 2, 1980 ■