Journal of Experimental Psychology:
Human Perception and Performance
1981, Vol. 7, No. 3, 604–610

0096-1523/81/0703-0604$00.75

Rapid Conceptual Identification of Sequentially Presented Pictures

Helene Intraub
Massachusetts Institute of Technology

When a sequence of pictures is presented at the rapid rate of 113 msec/picture, a viewer can detect a verbally specified target more than 60% of the time. In the present experiment sequences of pictures were presented at rates of 258, 172, and 114 msec/picture. A target was specified by name, by superordinate category, or by "negative" category (e.g., "the picture that is not of food"). Although the probability of detection decreased as cue specificity decreased, even in the most difficult condition (negative category cue at 114 msec/picture) 35% of the targets were detected. When the scores from the three detection tasks were compared with a control group's immediate recognition memory for the targets, immediate recognition memory was invariably lower than detection. The results are consistent with the hypothesis that rapidly presented pictures may be momentarily understood at the time of viewing and then quickly forgotten.

Recognition memory for pictures is remarkably good. Hundreds and even thousands of complex pictures that are presented for a few seconds each can later be recognized with better than 90% accuracy (Nickerson, 1965; Shepard, 1967; Standing, Conezio, & Haber, 1970). High levels of recognition memory performance are also obtained when pictures are presented for only about 100 msec each, as long as they are separated by substantial interstimulus intervals (ISIs). This occurs not only when the ISI contains a blank field (e.g., Rosenblood & Pulton, 1975) but also when it contains a familiar picture that repeats throughout the sequence (Intraub, 1980). If the ISIs are eliminated, however, and pictures are presented at the average rate of eye fixation (333 msec/picture) and faster (up to about 125 msec/picture), the normally excellent level of recognition memory obtained for pictures suffers dramatically, approaching the level of chance (Potter & Levy, 1969).

It is interesting that recognition memory should decline so precipitously when presentation rate mimics the rate at which visual scenes are normally fixated. Surely the visual system is not constructed so that the observer will scan the environment faster than each glimpsed scene can be analyzed. In normal vision, however, unlike the pictorial sequences used in those studies, there is considerable visual and conceptual overlap among successively fixated scenes. It has been proposed that this continuity allows expectancies to build up that serve to guide and to facilitate perception (e.g., Neisser, 1976). Based on this viewpoint it could be argued that the poor recognition memory performance obtained following rapid continuous presentation of pictures results from the observer's inability to identify the unrelated pictures. Potter (1975, 1976) has argued against this hypothesis, suggesting instead that in spite of the lack of continuity, virtually all pictures in such sequences are *momentarily* identified but then immediately forgotten. The locus of interference is placed not on identification but on the encoding processes necessary for retention.

In Potter's (1976) experiment sequences of magazine photographs were presented at rates ranging from 113 to 333 msec per pic-

This research was supported by Advanced Research Projects Agency Contract MDA 903-76-C-0441 to Massachusetts Institute of Technology. A preliminary version of this article was presented at the Eastern Psychological Association Convention, Philadelphia, April 1979. The author would like to thank Mary C. Potter, Ronald Finke, Joan M. Ryder, Nancy Woodard Cain, and David Milne for their extremely helpful comments and criticisms during preparation of the manuscript. The author also wishes to thank Michael Abrams for his assistance in data collection and scoring.

Requests for reprints should by sent to Helene Intraub, who is now at the Department of Psychology, University of Delaware, Newark, Delaware 19711.

ture. To determine whether pictures are momentarily identified, the ability to detect a cued picture in a sequence was compared with a control group's recognition memory for pictures presented at the same rate. Cuing in the detection task 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"). Consistent with the hypothesis that each glimpsed picture is independently identified and understood, although not necessarily retained, detection accuracy far surpassed recognition memory at all rates, regardless of the type of cue employed. At the most rapid rate (which was considerably faster than the average rate of eye fixation), 64% of the target pictures were detected with a verbal cue, whereas only 11% of the pictures were remembered. The proportion of pictures detected was interpreted as reflecting the minimal proportion of pictures momentarily identified.[1] This interpretation implies that identification of successive visual scenes is possible even without the continuity and expectancy characteristic of normal viewing.

Expectancy may not have been eliminated in Potter's (1975, 1976) experiments, however, because the verbal cue used in the detection task may have raised probabilistic anticipations about the visual attributes of the target. The high level of detection accuracy may simply reflect the fact that the cued picture was perceived more often than other pictures in the sequence. Similarly, it has been suggested that the "higher order" conceptual information provided by the verbal cues in Potter's experiments served to guide and facilitate the processing of "lower order" information, such as specific object identity (Carr & Bacharach, 1976; Neisser, 1976). If this is the case, then support for Potter's momentary identification hypothesis would be reduced to an artifact.

One way to determine whether expectancy alone causes detection accuracy to surpass recognition memory is to employ a detection task in which a picture is cued *without* giving the observer any specific information about its physical or conceptual characteristics, thereby eliminating perceptual priming. In the present experiment two different detection tasks using nonspecific cues were employed to this end: negative cuing and category cuing. Sequences were constructed that consisted of a diverse set of pictures belonging to a single general category and one picture that was not a member of that category. Subjects in the negative cue condition were provided with the name of the general category of the sequence and were instructed to detect and describe the picture that did *not* belong. In principle, to detect the negatively cued target picture, the subject would have to momentarily identify and categorize all pictures up to and including the target.

Although a negative cue should effectively limit any effect of expectancy there was some concern about the subject's ability to carry out the task for another reason. A negative decision is a relatively lengthy and difficult decision to make (e.g., Just & Carpenter, 1971). Under the time pressures of rapid presentation, a negatively phrased verbal cue could reduce detection ability for this reason alone. To circumvent this problem a positive detection task using a nonspecific cue was also employed. This constituted the category cue condition. Detection was cued in this case by providing the subject with the name of the superordinate category to which the target picture belonged. A category cue, though providing more information about the target than a negative cue, provides little advance information about the visual characteristics or identity of the target. In a third detection task, the target was cued by its specific name. This task, which is comparable to Potter's verbal cue task, was included as a replication.

A control group's immediate recognition memory performance was used for comparison with each of the three detection tasks. Because Potter's (1976) momentary identification hypothesis states that more pictures are identified than are remembered, it is important to avoid using any test procedures that might artifactually lower recognition

[1] "Identification" refers to the point at which the main theme of the picture is correctly determined. This does not imply covert naming of the pictures (which would be unlikely at the rapid rates of presentation employed), since pictures can be conceptually understood before they are named (Potter & Faulconer, 1975).

memory performance when testing this hypothesis. To make the recognition test as sensitive to memory for the target as possible, the following three provisions were introduced: (a) Unlike Potter's recognition test in which all of the pictures from a sequence were tested (yielding a 32-item test, including the distractors), in the present experiment only the target picture and one other picture from the sequence were tested. This procedure eliminated the interference that a series of relatively long tests might provide. (b) To enhance recognition memory the two distractors (new pictures) used in the brief 4-item test were neither visually nor conceptually similar to the target. (c) Furthermore, to make a more precise comparison between detection and recognition memory than in Potter's experiments, detection accuracy was compared with recognition memory for the target itself, rather than being compared with overall recognition memory performance.

If detection accuracy surpassed recognition memory in Potter's (1975, 1976) experiments because of expectancy, then the difference between detection and memory should be eliminated when nonspecific detection cues are used (i.e., in the negative cue and category cue conditions). If detection superiority is maintained in the nonspecific cue conditions, then this would indicate a striking ability of the observer to *momentarily* analyze and understand the contents of successively glimpsed scenes.

Method

Subjects

Subjects were 96 Massachusetts Institute of Technology undergraduates reporting normal or normal-corrected vision.

Materials

The stimuli were color magazine photographs of objects and scenes used by Intraub (1979, 1980). Eleven sequences of 12 pictures each were photographed, two frames per picture, using 16-mm movie film. Eleven pictures in each sequence belonged to a single general category (transportation, house furnishings and decorations, mechanical devices, food, body parts, people, animals, fruits and vegetables, and household appliances and utensils). The 12th was from a different category and appeared in a serial position between 2 and 11. This picture will be referred to as the target. Pictures within each general category were selected so that they would be as visually dissimilar as possible. For example, pictures of animals included creatures as diverse as a frog, a giraffe, a butterfly, and a dog; pictures of house furnishings and decorations included such items as a chandelier, pillows, and a chair. The target did not differ distinctively in size or in overall coloration from other pictures in the sequence. Two sets of 11 sequences were made, identical except that a different collection of targets appeared in each. Each set was presented to half of the subjects in every condition, so 20 different target pictures were tested over the course of the experiment. Slides of the pictures (35 mm) were used in the recognition test.

Apparatus

Sequences were projected on a screen using an L-W variable speed 16-mm movie projector. To obtain reaction times in the detection tasks, a bright white square was photographed in the lower corner of a frame, eight frames prior to the first picture in the sequence. The projected image of the square illuminated a photocell that triggered a digital reaction timer. The timer was stopped when the subject pressed a response button. A Kodak Carousel 35-mm slide projector was used to project targets and distractors in the recognition test. Size and illumination were approximately the same in the recognition test as in the inspection series. Pictures varied slightly in size; on the average the visual angle subtended by a picture was approximately 5° × 5°.

Design and Procedure

The three detection tasks and the recognition memory task were all tested at each of three presentation rates—114, 172, and 258 msec per picture—using a between-subjects design. Eight subjects were assigned to each of the 12 conditions. Subjects were presented with 1 sample sequence to familiarize them with the task and 10 experimental sequences. For half of the subjects in each condition, the sequences were run in reverse; thus each target appeared in two different serial positions. Film direction and target set were counterbalanced in each condition.

Target detection. Subjects in the detection groups were provided with the target cue prior to the start of each sequence. They were instructed to press the response button as soon as they saw the cued picture and to describe it briefly. Detection of the target was cued by specific name, by superordinate category, or by negative category. For example, if the general category of the sequence was house furnishings and decorations and the target was a picture of a butterfly, subjects in the specific name cue condition were told to "look for a butterfly," in the category cue condition they were told to "look for an animal," and in the negative cue condition they were told to "look for a picture that is *not* of house furnishings and decorations."

Recognition memory. Subjects in the recognition memory groups were instructed to attend to each picture in the sequence and to try to remember as many as possible. For this group no mention of categories or odd pictures was made. Following presentation of each sequence a four-item serial yes–no recognition test was immediately administered. The two old pictures were always the target picture and one nontarget from the sequence. Nontargets preceding and following the target in the sequence were tested equally often (the pictures immediately preceding and following the target in the sequence were not tested). Two new pictures (distractors) were included in each test. One distractor (the similar distractor) was a picture belonging to the general category of the sequence. The other distractor (the dissimilar distractor) was a picture that belonged neither to the general category of the sequence nor to the same category as the target. This testing procedure allowed for maximum test sensitivity for recognition of the target and allowed for assessment of guessing strategies.

Although no mention was made of categories or odd pictures in the recognition memory condition, the comments of a pilot subject indicated that subjects might in fact spontaneously notice the category-plus-odd-picture arrangement of the sequences. To determine whether this was the case, at the end of the session each subject in the recognition memory condition was asked to write a general description of the sequences shown in the experiment.

Scoring. In Potter's (1975, 1976) experiments the subject pressed a response key to indicate detection of the target in the sequence. A response was considered correct if it occurred between 250 and 900 msec following target onset. Because presentation was continuous this leaves some uncertainty about whether the response was indeed made to the target or if it was made to another picture. The problem was eliminated in this experiment by requiring the subject to describe the target. The description had to contain information specific enough to assure that the subject had identified the target in order for the response to be counted.

In the specific name cue condition, subjects were required to provide some specific information about the visual attributes of the target. For example, consider the target cue "chair." The chair in question was a reddish-brown leather easy chair with buttons on the back support. Responses such as "upholstered," "easy chair," "arm chair," "leather with buttons," "reddish-brown," or "side view with back support on right" were all considered correct. If the only description offered was a property inherent to all chairs (e.g., it had a back), the response was not counted as correct. (This type of response was rare.) In the category and negative cue conditions, where the specific identity of the object had not been provided, either a specific name or a description was counted as correct. For example, if the target was a brown shaggy dog, the responses "dog" and "shaggy brown animal" were accepted.

Table 1
The Proportion of Pictures Detected by Name, Category (Cat), and Negative Category (Neg) and the Proportion of Target Pictures and Nontarget Pictures Recognized at Each Rate of Presentation

Rate (msec/picture)	Detection			Recognition memory[a]	
	Name	Cat	Neg	Target	Nontarget
258	.89	.69	.79	.58	.58
172	.86	.71	.58	.49	.33
114	.71	.46	.35	.19	.34

[a] Recognition scores were corrected for guessing (see Footnote 2).

Table 2
Mean Reaction Time for Detection by Name, Category (Cat), and Negative Category (Neg) at Each Rate of Presentation

Rate (msec/picture)	Reaction time (in msec)					
	Name		Cat		Neg	
	M	*SD*	*M*	*SD*	*M*	*SD*
258	467	38	615	116	619	72
172	485	52	595	55	728	186
114	516	116	670	259	786	258

Results

Detection accuracy and recognition memory performance will be reported separately and then compared.

Detection

Subjects were able to detect and describe target pictures at all three presentation rates, even when they were provided with only a negative cue. The proportion of targets detected using name, category, and negative cues at each rate is shown in Table 1. Reaction times for those responses are shown in Table 2.

A two-way analysis of variance for Cue Type × Presentation Rate revealed an improvement in detection ability when slower presentation rates were used, $F(2, 63) = 13.39$, $MS_e = .0314$, $p < .001$, and when more specific cues were used, $F(2, 63) = 16.11$, $MS_e = .0314$, $p < .001$. A significant interaction between rate and cue type was not obtained, $F(4, 63) = 1.60$, although the interaction approached significance in a trend analysis, $F(2, 63) = 2.66$, $p < .10$. A similar analysis of the reaction times revealed that subjects responded more quickly as more specific cues were used, $F(2, 63) = 13.00$, $MS_e = 23{,}056.59$, $p < .001$. Unlike the accuracy scores, however, reaction time was unaffected by changes in presentation rate. An interaction was not obtained in either the analysis of variance or a trend analysis ($F < 1$, both cases).

Most detection failures were misses: .16, .32, and .35, for the name, category, and negative detection tasks, respectively. The proportion of trials that were counted as failures because of erroneous or nonspecific descriptions was .02, .06, and .08, respectively.[2] In no case did a subject report detecting a target without being able to provide a description; that is, no subject said, "I saw something that was *not* an animal, but I don't know what it was or what it looked like."

Recognition Memory

The mean proportion of target pictures and the mean proportion of nontarget pictures recognized at each rate are shown in Table 1, corrected for guessing.[3] A two-way analysis of variance for Picture Type (target or nontarget) × Presentation Rate, with repeated measures on picture type, was conducted. As expected, fewer pictures were recognized as faster presentation rates were used, $F(2, 21) = 6.19$, $MS_e = .3963$, $p < .001$. No difference in the corrected proportion recognized was obtained between target and nontarget pictures ($F < 1$), nor was there an interaction between picture type and presentation rate ($F < 1$). Nontargets preceding and following the target picture in the sequence were remembered equally often.

There were virtually no false yeses (.02) made to the dissimilar distractor (the target picture's control for guessing). In fact, only one subject committed this type of error. This shows that subjects did not simply respond "yes" to any test picture that did not belong to the general category of the sequence. As would be expected following rapid presentation of 11 pictures from the same category, false yeses to the similar distractor (the nontarget picture's control for guessing) were relatively abundant (.25).

At the end of the session, when the subjects were asked to describe the sequences they had just seen, all subjects reported that pictures were grouped by category and 83% specifically reported a category-plus-odd-picture arrangement.

Detection Versus Recognition Memory

Contrary to the expectancy hypothesis, more pictures were detected during presentation than were remembered immediately following presentation regardless of the specificity of the detection cue (see Table 1), thus providing strong support for the momentary identification hypothesis. Detection accuracy for each cue type was individually compared with recognition accuracy for the target picture, collapsing over rate. Planned comparisons revealed that in each case, significantly more pictures were detected than were remembered; $F(1, 84) = 64.00, 16.00$, and 9.00, $MS_e = .0334$, $p < .001$, when recognition memory was compared with detection by name, by category, and by negative category, respectively. A two-way analysis of variance for Task × Presentation Rate, including all four tasks, revealed no interaction between task and rate, $F(6, 84) = 1.21$. The main effects of task and presentation rate were both highly significant, $F(3, 84) = 19.97$, $MS_e = .6676$, $p < .001$; $F(2, 84) = 24.39$, $MS_e = .8154$, $p < .001$, respectively.

---

[2] Some of the erroneous responses in the category and negative detection conditions nonetheless might reflect analysis of the target. An example of this type of response is describing a long-legged dog as a deer. Particularly in the negative cue condition, where the target could be virtually anything, the response *deer* would seem to indicate some analysis of the target–like the target, a deer is a brownish four-legged creature. When more lenient scoring was employed, category and negative detection scores increased slightly. For the slow to fast rates, respectively, the category detection scores were .70, .72, and .48, and the negative detection scores were .80, .63, and .40.

[3] The formula used to correct for guessing was $Y_c = (TY - FY)/(1 - FY)$, in which Y_c is the corrected proportion of *yes* responses, TY is the proportion of correct *yes* responses to old pictures, and FY is the proportion of *yes* responses to distractors (false yeses). The proportions of negative pictures and general category pictures recognized were corrected separately using the appropriate distractor (i.e., the dissimilar distractor was employed in the target picture correction, and the similar distractor was employed in the nontarget picture correction).

Discussion

The results show that observers possess a striking ability to identify and understand unrelated pictures at presentation rates equal to or faster than the average rate of eye fixation. Detection of verbally cued pictures was superior to a control group's immediate recognition memory for the same pictures, even when a highly sensitive test of recognition memory was employed. This supports Potter's (1976) hypothesis that pictures are momentarily understood but immediately forgotten during rapid presentation.

Furthermore, the results clearly demonstrate that expectancy alone cannot account for the superiority of detection ability over recognition memory. Detection was superior when the target had been negatively cued, that is, even when no specific information regarding the visual or conceptual characteristics of the target was provided (e.g., "detect and describe a picture that is *not* of an animal"). In fact, at the rate of presentation that most closely approximates the average rate of eye fixation (258 msec/picture), 79% of all targets were detected and described on the basis of a negative cue, whereas only 58% of all target pictures were remembered immediately following presentation.

Additional support for the momentary identification hypothesis is provided by the recognition memory subjects' descriptions of the sequences. These subjects were not given any hint that the sequences contained diverse pictures from a general category and one picture that did not belong. They were simply instructed to pay attention to each picture in the sequence and to try to remember them all. Yet when asked to describe the sequences at the end of the session, all subjects reported that pictures were grouped by category, and 83% wrote specifically that the sequences were arranged in a category-plus-odd-picture fashion. Although their immediate recognition memory was relatively poor, apparently these subjects had momentarily identified and categorized the pictures during presentation. In their descriptions a few subjects indicated that they could have adopted a strategy to remember the odd picture, but that doing so would have disrupted their ability to remember the other pictures in the sequence. The observation that nontargets preceding and following the target were remembered equally often suggests that consistent with those reports, subjects did not adopt that strategy.

One might still argue that expectancy played some role in detection because, overall, detection accuracy increased as more specific cues were used. According to this view expectancy may have increasingly facilitated perception of the target as more specific cues were provided. There is, however, an alternative explanation of the difference among the three detection tasks. The process of deciding that the cue and the target picture match increases in complexity as less specific cues are provided.

Consistent with this consideration, responses were fastest when the target was cued by specific name and slowest when it was cued by negative category. If pictures are identified only momentarily during presentation, then a decision process requiring relatively little time is more likely to be completed before attention is drawn to the next pictures. Thus detection by specific name is more likely to be successfully completed than detection by category or by negative cue. In fact, when the time per picture was increased to 258 msec, detection by negative cue was almost as accurate as detection by specific name, yet the response required an average of 150 msec longer.

The results do not imply that expectancy is not important in normal visual perception. What they do show is that even without the continuity characteristic of normal vision, successively glimpsed scenes can be understood surprisingly well. This ability may function as a monitoring system in normal vision. For example, momentary identification of each fixated scene may play a role in controlling placement of subsequent eye fixations. (For an example in reading, see Rayner, 1979.) Establishment of a relatively stable memory representation of a scene requires more than identification (contrary to earlier suggestions; cf. Haber, 1970). Storage requires implementation of encoding processes that can extend beyond the duration of the stimulus (Intraub, 1979, 1980; Potter, 1976; Tversky & Sherman, 1975;

Weaver, 1974; Weaver & Stanny, 1978). During rapid presentation of unrelated scenes, when the observer attempts to analyze and remember them all, these encoding processes are probably disrupted when attention shifts from one picture to the next picture in the sequence (Intraub, 1980, 1981; Potter, 1976; Potter & Levy, 1969). At extremely rapid presentation rates apparently the primary locus of interference is not at the level of conceptual identification but at the level of memory.

References

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

Haber, R. N. How we remember what we see. *Scientific American*, 1970, *222*(5), 104–112.

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

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.

Intraub, H. Identification and processing of briefly glimpsed visual scenes. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eyemovements: Cognition and visual perception*. Hillsdale, N.J.: Erlbaum, 1981.

Just, M. A. & Carpenter, P. A. Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 244–253.

Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.

Nickerson, R. S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, 1965, *19*, 155–160.

Potter, M. C. Meaning in visual search. *Science*, 1975, *187*, 965–966.

Potter, M. C. Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 509–522.

Potter, M. C., & Faulconer, B. A. Time to understand pictures and words. *Nature*, 1975, *253*, 437–438.

Potter, M. C., & Levy, E. I. Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 1969, *81*, 10–15.

Rayner, K. Eye guidance in reading: Fixation locations within words. *Perception*, 1979, *8*, 21–30.

Rosenblood, L. K., & Pulton, T. W. Recognition after tachistoscopic presentations of complex pictorial stimuli. *Canadian Journal of Psychology*, 1975, *29*, 195–200.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 156–163.

Standing, L., Conezio, J., & Haber, R. N. Perception and memory for pictures: Single-trial learning of 2560 visual stimuli. *Psychonomic Science*, 1970, *19*, 73–74.

Tversky, B., & Sherman, T. Picture memory improves with longer on-time and off-time. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *1*, 114–118.

Weaver, G. E. Effects of poststimulus study time on recognition of pictures. *Journal of Experimental Psychology*, 1974, *103*, 799–801.

Weaver, G. E., & Stanny, C. J. Short-term retention of pictorial stimuli as assessed by a probe recognition technique. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 55–65.

Received July 26, 1979
Revision received May 12, 1980 ■