

Changes in picture recognition memory over time using an exclusion set paradigm

J. ELIZABETH BIRD

The Australian National University, Canberra, Australia

Previous picture recognition studies have been insensitive because of ceiling effects (e.g., Standing, 1973) and have not dealt adequately with the problem of partial recognition. A recognition measure was devised to determine the amount of detail in picture memory on a relative scale. Complex line drawings were successively altered to make stimulus sets consisting of a prototype for presentation and a set of 15 cumulative variations (distractors) which were shown serially to subjects during recognition testing and rated for the confidence with which they could be rejected. The size of the set of pictures subjects could exclude was scored. Testing was done immediately or after delay periods of 1, 2, 7, or 60 days. Exclusion set size decreased significantly over delay.

Studies comparing recognition memory for pictures and words (e.g., Bird & Bennett, 1974; Jenkins, Neale, & Deno, 1967; Paivio & Csapo, 1969; Shepard, 1967) suggested that picture memory has a phenomenal capacity. The presence of ceiling effects in picture recognition (Nickerson, 1965; Shepard, 1967) even over long periods of time (Nelson, Metzler, & Reed, 1974; Nickerson, 1968) or for vast sets of stimuli (Standing, 1973) has led to the idea that picture memory is almost photographic and may not require elaborate encoding or retrieval strategies. However, the ceiling effects may merely indicate that present picture memory paradigms are insensitive and overestimate memory capacity. More sensitive recognition studies have demonstrated limits on the capacity of picture memory. These studies have used "homogeneous" distractor sets (Bairick & Bairick, 1971; Bairick & Boucher, 1968; Bairick, Clark, & Bairick, 1967; Goldstein & Chance, 1971; Howe, 1967), where the test item and its distractors can be described by the same verbal label, a concrete noun. No studies have tackled the most pressing problem in picture recognition, that of "partial recognition" (see Postman, 1963) where the subject is able to use only partial information about the stimulus in order to discriminate between it and its distractors, thus spuriously inflating memory scores. In experiments where the stimulus differs from its distractors in many ways, any one of these could act as a recognition cue.

Developmental studies have used more homogeneous stimuli and distractors, where distractors are transformations of the original stimulus (Brown & Campione, 1972; Mandler & Stein, 1974). This decreases the probability that partial recognition is an important factor in recognition, although each distractor may still be different in many ways from the target item.

Requests for reprints should be sent to J. Elizabeth Bird, Psychology Department, The Australian National University, Canberra, A.C.T., Australia 2600. This paper is sponsored by P. D. McCormack, who takes full editorial responsibility for it.

In the present study a new method of recognition was developed where a prototype picture was used as the stimulus and 15 recognition distractors were made by altering the prototype cumulatively 15 times. The prototype and its first variation differed only in some small detail, the second variation differed from the first variation in one detail and from the prototype in two details, and so on, so that each successive variation from the prototype was less similar to it. It was assumed that the details of the prototypes chose to be changed were a random sample of possible changes to the pictures and that subjects' loss of details of the pictures would be idiosyncratic. Provision for the latter assumption was found in a study of recall of the pictures by drawing, which indicated the details most salient to subjects which could be avoided in making changes to the prototypes. Hence spuriously high recognition scores due to partial recognition could be made negligible.

The rationale for this method was that subjects' memory for pictures is not perfect and that for every subject attempting to remember a prototype there would be some point on the continuum of recognition items (from prototype to the 15th variation) above which the subject would be able to reject all variations as clashing with his memory representation and below which the subject would be unable to discriminate between the prototype and its closest variations. This view of recognition suggests that at perception subjects construct a description of the stimulus, which contains information which may be unavailable for retrieval later. In the recognition task the subject may match this memory description with the percept of the recognition test item presented to him. The subjects would be able to reject any pictures which had inconsistent features with their memory reconstructions but would not be able to reject the pictures which were consistent with their "faded" descriptions of the original. The present recognition method allows for 15 different levels of

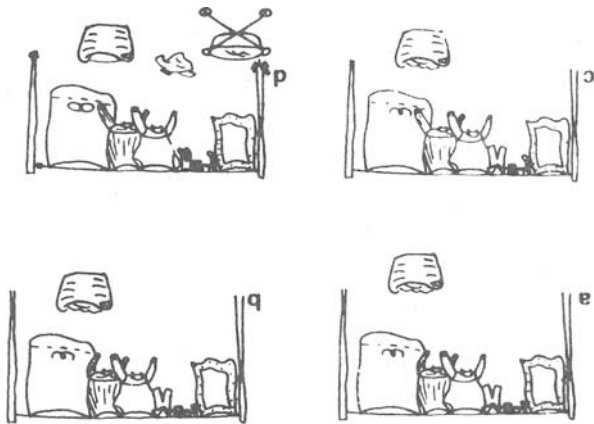


Figure 1. (a) One of the prototype stimuli. (b) In variation 1 the laundry basket is closer to the clothesline. (c) Variation 7. (d) Variation 15.

memory description to be assessed.

This method was used to investigate changes in picture memory over time. Several studies have shown some loss of picture memory over delays of up to 1 year, although they do not report whether these differences are significant (Bahrick & Boucher, 1968; Nickerson, 1968; Shepard, 1967; Standing, Conezio, & Haber, 1973). Significant changes have been reported for changes over 2 days (Goldstein & Chance, 1971) and over 7 weeks (Nelson et al., 1974). Pilot testing with the present method suggested that changes may occur after a few days and drop at a faster level over time than has been found in previous studies. Five delay groups were chosen, the subjects had immediate testing (no delay group) or delays of 1 day, 2 days, 7 days or approximately 2 months (60 days).

METHOD

Subjects

Seventy first-year psychology students were assigned randomly to each of the five delay groups. There were 7 females and 7 males in each group.

Stimuli

Six complex, black and white line drawings were chosen from a set of 20 pictures adapted from children's books. Pictures were chosen on the basis of pilot testing; pictures with large ranges in recognition scores or which appeared to be too easy or too difficult were discarded. Each picture contained approximately 10 items in an integrated scene.

Each prototype drawing was altered 15 times to create a set of distractor items. Alterations were made by adding or deleting lines, changing shapes of items or moving whole items to different positions. The changes were cumulative and made a stimulus continuum of 16 items, all on a linear scale of amount of change from the prototype so that Variation 15 was least like the original. The amount of line change in each variation was not the same; early variations involved minimal line changes, while later variations involved more line change. A group of judges determined that changes made to the pictures were not unidimensional and that it was impossible to scale variations for the perceived similarity between them. Scaling the objective amount of line change in each variation was avoided, as this

appears to be unrelated to the subjects' perceptions of the amount of change in each variation (cf. stimuli used by Daniel, 1972). A prototype and three of its variations are shown in Figure 1. The prototypes and variations were reduced to 35-mm slides. There were four test items, one practice item (the task was easily understood after one test trial), and a "baseline item" used to gauge the memory abilities of the subjects in different groups.

Apparatus

Two Kodak Carousel slide projectors, Model S-RA, were used to project pairs of pictures side by side onto a rear projection screen in front of the subject. Each slide projected onto an area 26 x 17 cm, 10 cm apart, with the subject 75 cm away from the center of the screen. The experimenter sat by the subject at a table with the controls for the projectors and an electronic timer which operated a black metal shutter to cover the right projector's lens at timed intervals. Polaroid filters were placed in front of each lens to lessen glare.

Procedure

Subjects were tested individually. They were first shown a practice item and told to look at it closely because they would later be shown "a whole series of very similar pictures to see how well (they) can recognize the original picture among a lot of similar ones." The practice item was presented for 10 sec and then tested for recognition with its stimulus distractor set. The same procedure was carried out with the baseline item. The four test prototypes were then shown successively, with a 10-sec presentation time and a 10-sec interstimulus interval. The four test pictures were randomly ordered for each subject. Test pictures were presented on the right side of the screen.

Delayed-group subjects were given the practice and baseline items and tested for recognition of these; they were then presented with the four test stimuli "to remember for next time." Upon returning for recognition testing following the appropriate delay, these subjects were again given the initial instructions and the recognition sets for the four test items. No-delay subjects were given the entire procedure in one sitting. Recognition testing was done in two phases.

Phase 1. The subjects were presented with the 16 pictures in a set in an individually randomized order. They made 7-point ratings about each picture from "1," positive that the picture is the original, to "7," positive that the picture is not the original. The subjects were told to make four responses (not sure) instead of guessing. The ratings were a partial "blind," used to minimize guessing correctly. Because of the way the sets were organized, when subjects correctly rejected a distractor with a 7 rating, they were not shown variations higher on the continuum with redundant line changes. If such subjects failed to give 7 ratings to higher variations, it would either be because they failed to notice the mistake in the picture which they had noticed before or because they changed their memory of the item. It was assumed that subjects did not change memory storage during the course of testing and that refusals to reject high variations would be because of an attentional failure in looking at the picture during testing rather than a memory failure. Since 7 ratings were important in deriving recognition scores, the subjects were asked to give their reasons for making rejections. Pilot testing had shown that subjects often make incorrect rejections, so only correct rejections were counted in scoring.

Phase 2. Because it was thought that ordinary recognition procedures may not distinguish between attention failures and memory failures during testing, after Phase 1 subjects were shown pairs of pictures in the set to see if they could reject any more items given a second chance. Of the remaining items on the stimulus continuum which the subjects had not correctly rejected, a randomly chosen picture was again presented to the subject, along with its neighbor above or below it on the stimulus continuum. The difference between these two items was pointed out to the subject and they were asked if they could

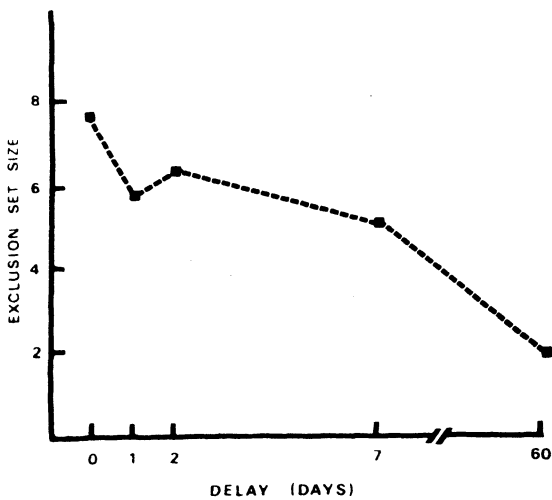


Figure 2. Recognition memory (exclusion set size) as a function of delay.

reject either item on the basis of this difference. Subjects either gave four responses or rejection ratings of 5, 6, or 7 to one of the pictures. If a picture was correctly rejected with a 7 rating, no higher variations from the prototype were presented.

Scoring. The measure of recognition accuracy used in the present experiment was that of "exclusion set," or the number of items given 7 rejections. Since some of these items were not directly presented to subjects, the score was actually 17 minus the number of the lowest variation given a correct 7 rating.

RESULTS

To test whether the different delay groups differed in memory ability, a one-way analysis of variance was carried out on the exclusion set scores for the baseline item. Means for the groups were: no delay-7.857; 1 day-7.214; 2 days-8.071; 7 days-7.643; 60 days-7.714. There was no significant effect of delay group, $F(4,60) < 1$, indicating that groups did not differ in initial ability. Because of this it was not considered necessary to carry out an analysis of covariance on test scores.

The baseline scores were also subjected to an analysis by sex. Scores for all delay groups were pooled so that a t test comparing means on baseline items for all female subjects and all male subjects could be carried out. No significant difference was found, $t(.025, 68) < 1$.

Mean scores for each delay group over all pictures are shown in Figure 2. A two-way analysis of variance with repeated measures on the picture factor was carried out on the data. There was a significant effect of delay, $F(4,65) = 9.661, p < .01$, but not for pictures, $F(3,195) = 2.470$, or for the interaction of pictures with subjects within groups, $F(12, 195) = 1.468$. Thus there was a significant change in memory for pictures over time. There was no difference between the four pictures.

Scheffé comparison were carried out after the analysis of variance to test which difference between means of

delay groups were significantly different. The no delay group differed significantly from the 7-day group ($p < .1$) and the 60-day group ($p < .01$) but not from the 1-day and 2-day groups. Both the 1-day and 2-day groups were significantly different from the 60-day group ($p < .05$ and $p < .01$, respectively), but they did not differ significantly from each other. The 7-day group differed significantly from the 60-day group ($p < .1$). The 10% level of significance was used as the criterion for acceptance, as Scheffé (1954) recommended due to the conservatism of his measure.

DISCUSSION

A new picture recognition paradigm was devised, which was sensitive to memory changes over time. By sampling many different changes in a single picture during recognition testing, it was found that picture memory does not have the enormous capacity suggested by previous studies (Standing, 1973). Picture memory dropped in accuracy from immediate testing to 7 days and again after 2 months, but not after a shorter period, as suggested by Goldstein and Chance (1971) for recognition of inkblots and snowflakes. It may be that decay functions for picture memory are not the same for different types of stimuli. It would be extremely difficult to map a true decay function for picture memory, since it is impossible to devise any meaningful free recall for visual detail. Recognition testing is limited by the necessity of providing subjects with distractors which they must eliminate. Difficulty of a recognition task is a direct function of the difficulty of its distractor sets. Also, there is the problem that the presence of distractors may interfere with or even distort the subjects' memories for the stimuli during testing. Research with the exclusion set method should be directed towards minimizing or controlling these difficulties. The exclusion set method has many advantages over previous methods. For example, 91% of the subjects improved their recognition scores during Phase 2 of testing. This suggests that ordinary recognition experiments cannot estimate memory adequately if subjects are responding incorrectly due to attentional failures during testing.

The difficulty of the present recognition task was illustrated by the presence of low exclusion set scores and the subjects' comments in postrecognition interviews. Recognition may not be an automatic process lacking elaborate retrieval strategies, as has been suggested (Kintsch, 1970; McCormack, 1972). There is no evidence to assume that picture recognition does not require retrieval processes. It may be that recognition requires the same kind of retrieval as free recall. Hopefully, future research will overcome the technical limitations of present recognition methods which do not make it feasible to compare retrieval processes for different types of stimuli or different memory paradigms.

There was some evidence that subjects can encode pictures verbally or by some other means whereby little visual detail is stored. Three subjects in the 60-day group claimed never to have seen pictures like the stimuli during testing, though they admitted having seen pictures with the same types of items. This phenomenon makes long-term picture memory testing difficult unless heterogeneous distractors are used, and this type of task is plagued by the problem of partial recognition. It seems unlikely that subjects could rely only on verbal descriptions in their picture memories, since they are able to perform with accuracy on homogeneous recognition tasks. However, some verbal coding may serve to anchor the subjects's memories about the stimulus or to distort pictorial memory in favour of archetypal visual patterns (Carmichael, Hogan, & Walter, 1932; Daniel, 1972). With dual encoding verbal descriptions may decay less rapidly over time than memory for visual configurations.

Many subjects reported that they were able to exclude pictures which clashed with their memories of the prototype but they were uncertain about the remaining pictures. This was support for the view that subjects have some faded memory description of the stimulus which could give a recognition match to many distractor variations of the stimulus, as long as the perceptual description of each variation did not clash with the memory description. It may be a fruitful approach to regard perception of and memory for a picture as constructive activities (see Neisser, 1967) which lead to descriptions of the stimulus which can be compared in some internal matching process.

REFERENCES

- Bairick, H. P., Clark, S., & Bairick, P., Generalization gradients as indicants of learning and retention of a recognition task. *Journal of Experimental Psychology*, 1967, 75, 464-471.
- Bairick, H. P., & Bairick, P. Independence of verbal and visual codes of the same stimuli. *Journal of Experimental Psychology*, 1971, 91, 344-346.
- Bairick, H. P., & Boucher, B. Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 1968, 78, 417-422.
- Bird, J. E., & Bennett, A. F. A developmental study of recognition of pictures and nouns. *Journal of Experimental Child Psychology*, 1974, 18, 117-127.
- Brown, A. L., & Campione, J. C. Recognition memory for perceptually similar pictures in preschool children. *Journal of Experimental Psychology*, 1972, 95, 55-62.
- Carmichael, L., Hogan, H. P., & Walter, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, 1932, 15, 73-86.
- Daniel, T. C. Nature of effect of the verbal labels on recognition memory for form. *Journal of Experimental Psychology*, 1972, 96, 152-157.
- Goldstein, G., & Chance, J. E. Visual recognition memory for complex configurations. *Perception & Psychophysics*, 1971, 9, 237-241.
- Howe, M. J. A. Recognition memory for photographs in homogeneous sequences. *Perceptual and Motor Skills*, 1967, 24, 1181-1182.
- Jenkins, J. R., Neale, D. C., & Deno, S. L. Differential memory for picture and word stimuli. *Journal of Educational Psychology*, 1967, 58, 303-307.
- Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.
- McCormack, P. D. Recognition memory: How complex a retrieval system? *Canadian Journal of Psychology*, 1972, 26, 19-41.
- Mandler, J. M., & Stein, N. L. Recall and recognition of pictures by children as a function of organization and distractor similarity. *Journal of Experimental Psychology*, 1974, 102, 657-669.
- Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- Nelson, T. O., Metzler, J., & Reed, D. A. Role of details in the long-term recognition of pictures and verbal descriptions. *Journal of Experimental Psychology*, 1974, 102, 184-186.
- Nickerson, R. S. Short-term memory for complex meaningful visual configurations: a demonstration of capacity. *Canadian Journal of Psychology*, 1965, 19, 55-160.
- Nickerson, R. S. A note on long-term recognition memory for pictorial material. *Psychonomic Science*, 1968, 11, 58.
- Paivio, A., & Csapo, K. Concrete-image and verbal memory codes. *Journal of Experimental Psychology*, 1969, 80, 179-285.
- Postman, L. One-trial learning. In C. N. Cofer and B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes*. New York: McGraw-Hill, 1963, Pp. 295-321.
- Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 156-163.
- Standing, L. Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 1973, 25, 207-222.
- Standing, L., Conezion, J., & Haber, R. N. Perception and memory for pictures: single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 1970, 19, 73-74.

(Received for publication December 5, 1974.)