

Comprehension and memory for pictures

GORDON H. BOWER*, MARTIN B. KARLIN, and ALVIN DUECK†
Stanford University, Stanford, California 94305

The thesis advanced is that people remember nonsensical pictures much better if they comprehend what they are about. Two experiments supported this thesis. In the first, nonsensical "doodles" were studied by subjects with or without an accompanying verbal interpretation of the pictures. Free recall was much better for subjects receiving the interpretation during study. Also, a later recognition test showed that subjects receiving the interpretation rated as more similar to the original picture a distractor which was close to the prototype of the interpreted category. In Experiment II, subjects studied pairs of nonsensical pictures, with or without a linking interpretation provided. Subjects who heard a phrase identifying and interrelating the pictures of a pair showed greater associative recall and matching than subjects who received no interpretation. The results suggest that memory is aided whenever contextual cues arouse appropriate schemata into which the material to be learned can be fitted.

The following experiments address the question of how people remember pictures. We may begin with the observation that pictures (drawings, diagrams, photographs) comprise a two-dimensional notational system which, like language, has both a "surface structure" (the medium) and a meaningful "deep structure" (the message). Like language, pictures have a terminal vocabulary (of strokes, shadings, etc.), sets of combination rules, often a referential field, and conventional rules for interpreting what a picture is about (see Gombrich, 1960; Goodman, 1968). Pictures, especially "realistic" ones, denote objects or scenes in a manner that parallels the symbolic way that words and sentences do. And just as language appears to be acquired as a perceptual-motor skill, so also does it appear that children learn the conventional rules for interpreting the notational symbolism of pictures. These rules guide our construction of what a picture is about—what conceptualizations it expresses or what objects it symbolizes. That we learn to interpret drawings is illustrated clearly by the difficulty novices have acquiring the symbolic system of their profession, such as ballet labanotation, musical scoring, molecular structure, etc.

We are interested in memory for pictures. The hypothesis to be tested is that a major determinant of how well a person can remember a picture is whether or not he "understands" it at the time he studies it. If he comprehends the picture—achieves a compact interpretation of it—then he should remember it much better than if he fails to comprehend it.

This hypothesis was suggested by the work of Bransford and Johnson (1972), Bransford and McCarrell (in press), and Doll and Lapinski (1974) on memory for linguistic material. They showed convincingly that a

person's ability to recall a sentence depends on whether the sentence causes him to call to mind an appropriate referential situation. For example, consider causal sentences such as: (1) The notes were sour because the seams split; (2) The voyage wasn't delayed because the bottle shattered; (3) The haystack was important because the cloth ripped. Though simple in syntax and word meanings, such sentences prove difficult to understand and recall. The mind boggles because a causal connection is asserted to hold between two apparently unrelated events; the subject cannot call to mind an appropriate schemata (known scenario) into which the events can be substituted and thus related causally. But all difficulties dissolve if the subject is provided with a clue as to an appropriate causal schemata: The clue is a simple "thematic prompt" (for the three sentences above, *bagpipe*, *ship christening*, *parachutist*). The clue calls to mind a known scenario (see the "frames" theory of Minsky, 1974) into which the events mentioned in the sentence can then be fitted. The sentences then become comprehensible and memorable.

We wish to advance here a parallel argument for the role of comprehension in memory for pictures. Our experiments will, therefore, expose subjects to pictures which are very difficult to "understand" unless one is given a thematic clue; we then later test memory for pictures that had been shown with or without the clue. Of course, this means that we are investigating memory for "nonsensical" pictures, one for which subjects usually have no interpretation. But what makes a picture nonsensical or meaningless? There are doubtless several kinds of nonsense, but included would be pictures for which the viewer (1) does not know the conventions for interpretation (e.g., a musical score for a musician who only "plays by ear"), (2) does not know the conceptual denotations of the symbols (e.g., the step sequences corresponding to ballet labanotation), or (3) knows both of the above but still can achieve no coherent understanding by applying the standard conventions because the picture does not supply enough interpretive

*The authors thank Susan L. Karlin for creating most of the stimulus materials for Experiment II. The research was supported by Grant MH 13905-07 from the National Institute of Mental Health to the first author.

†Requests for reprints may be sent to Gordon H. Bower, Department of Psychology, Stanford University, Stanford, California 94305.

clues. Examples of the latter kind, which we shall use in our research, occur with "impoverished" pictures: These are pictures which reduce or eliminate the salient clues or distinctive features of objects which typically guide our selection of their schemata from memory. Such pictures present fragments of hidden figures which may be seen only by suggestion. They appear uninterpretable until a clue retrieves from memory an appropriate conceptual frame which can then be fit onto the line fragments.

A curious side effect results from finally finding a conceptual schema which fits: The tension of "What is it?" dissolves with laughter into "Oh, now I get it!" Many of us became familiar with such visual jokes in the early 1960s with the "doodles" rage in America: A doodle was an uninterpretable drawing that turned out to have a funny interpretation (see Price, 1972). Figures 1a and 1b show two of the examples used in our experiment.

EXPERIMENT I

The first experiment used free recall to assess the effects of comprehension on picture memory. Subjects studied a series of doodle pictures which they were to recall. Some subjects heard the interpretation as they saw each picture; other subjects, the controls, simply viewed the pictures without hearing any interpretive comment. The session ended with subjects drawing copies of those pictures they could recall.

A second, minor hypothesis tested was that subjects might distort their memory of the picture in a direction which provided a better fit to the prototype of the category used to interpret the picture upon original viewing. This "assimilation hypothesis" is an old one (see Carmichael, Hogan, & Walter, 1932; Riley, 1963), but the evidence regarding it has been equivocal. To test the hypothesis, we had subjects return after a week for a recognition memory test. Besides the correct picture, the multiple-choice set for each item contained two distractor pictures that were equally similar to the correct picture in terms of a line-overlap measure. One of these distractors exemplified a minor variation on the target which made it look even more like the interpreted prototype than did the original target (call this the "prototype" distractor). The other distractor involved a similarly minor line alteration but was done in such a manner as to violate the fit of the interpretive schema to the picture (call this the physically "similar" distractor). The expectation of the assimilation hypothesis is that, when recognition errors are made, subjects who learned the picture with a suggested interpretation will make relatively more errors on the prototype distractor than on the physically similar distractor. On the other hand, subjects who do not achieve the appropriate interpretation should tend to divide their errors evenly between the prototype and similar distractors.

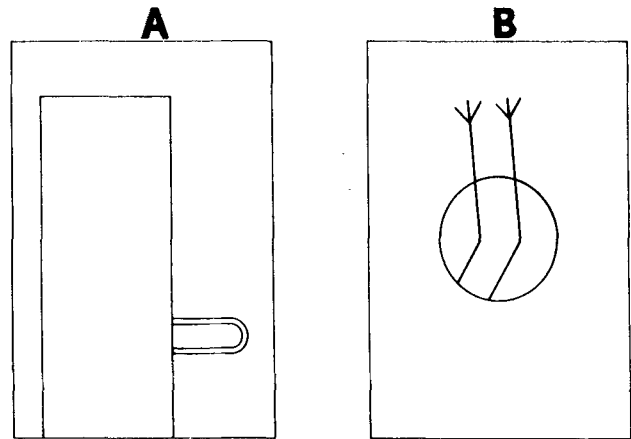


Figure 1. Doodles of Experiment I. Panel A: A midget playing a trombone in a telephone booth. Panel B: An early bird who caught a very strong worm.

Method

The subjects were 18 undergraduates fulfilling a service requirement for their introductory psychology course. They were tested individually, assigned in random order to the "label" or "no-label" conditions of the experiment. All subjects studied a series of 28 simple doodle pictures shown on 3 x 5 in. cards at a rate of one every 10 sec. As each picture was shown, its appropriate interpretation was given by the experimenter to the subjects in the label group but not to subjects in the no-label group. Following presentation of the list, subjects had 10 min to draw all the pictures they could remember in any order they wished. The recall sheets were 8 x 11 in. papers marked off into a 3 by 3 matrix; the subject was instructed to recall by quickly sketching a recalled picture in one of the nine boxes on his recall sheet and to use as many sheets as necessary to complete his list recall. Before recall commenced, it was emphasized that the subject should aim for sketching the "gist" of the pictures recalled rather than for providing a lot of artistic detail of each picture. (The pictures could in fact be drawn very simply.) Following completion of the recall task, the subject was dismissed with an appointment to return the next week "for other experiments."

Upon returning the next week, subjects received the three-alternative multiple-choice test over 24 of the 28 pictures of the originally learned list (for four of the original pictures, we were unable to think up two similar distractors which met our criteria). The subject received a six-page booklet, with four multiple-choice triplets arranged in rows down each page. He was told that each triplet (row) contained one picture he had seen the week before plus two closely similar pictures. He was asked to rank order the three alternatives in each row, placing a 1 beside that test figure he considered most like the one he remembered seeing, a 2 beside the next most similar one, and a 3 beside that picture he considered least similar to the one he remembered. The test was self-paced. Upon completing the test, the subject was debriefed and dismissed. One subject of the no-label group failed to return for the 1-week test, leaving eight subjects in that group at that point.

Results

Free Recall. A first noteworthy fact is that we had relatively few problems in scoring for "gist recall" of the sketches. We had anticipated severe problems produced

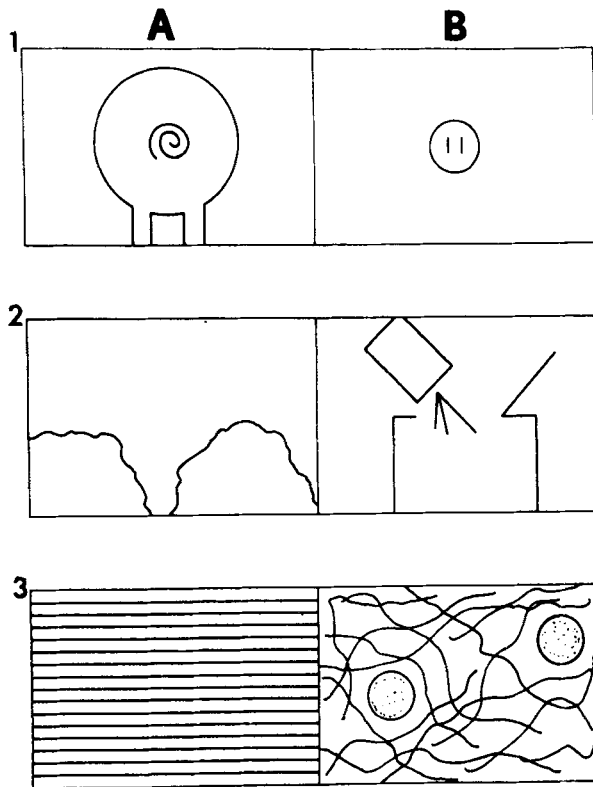


Figure 2. Pairs of nonsensical pictures used in Experiment II. See the text for an explanation of their contents.

by interfering or confused combinations of several pictures, or at least deletions causing the sketch to be unidentifiable. But subjects tended to recall (sketch) the pictures either relatively accurately or not at all. The primary result of interest is that an average of 19.6 pictures out of 28 (70%) were accurately recalled by the label group [standard error of the mean (SEM) = 1.25], whereas only 14.2 pictures (51%) were recalled by the no-label group (SEM = .92). The means differ reliably in the predicted direction [$t(16) = 3.43$, $p < .01$]. Thus, we have clear confirmation that "picture understanding" enhances picture recall.

Recognition Memory. Despite the closeness of the distractors to the target, recognition of the correct target at the 1-week retention interval was very high. Subjects who received labels during study correctly recognized (gave a 1 rating to) a mean of 22.0 out of the 24 test triplets (92%); subjects receiving no labels during study correctly recognized a mean of 20.1 pictures (84%). With standard errors of .83 and 1.08, respectively, the means do not differ reliably.

Even noting the high levels of recognition accuracy, we may still ask whether the label and no-label subjects react differently to the prototype vs. the physically similar distractor. There are several indications that the label subjects considered the prototype distractor much closer subjectively to the target. First, considering only

cases when the correct picture was not ranked first, the conditional probability that the prototype (rather than the similar distractor) was ranked first was .75 for the label subjects but only .38 for the no-label subjects. However, these conditional probabilities are based on very few observations. A more stable measure of differentiation is provided by the difference in rankings (on similarity to the remembered target) between the prototype distractor and the physically similar distractor. For the label group the mean rank assigned to the prototype was 2.17, whereas that for the similar distractor was 2.76, a difference of .59. In contrast, for the no-label group, the rankings of the two distractors was closer: 2.34 for the prototype and 2.48 for the similar distractor, a difference of only .14. The difference in rankings is reliably larger for the label group than for the no-label group [$t(15) = 2.79$, $p < .02$]. This result accords with the assimilation hypothesis: Subjects receiving the picture interpretation during study later reported that the distractor which moved in the direction of the interpreted prototype was closer to the target than was the distractor which involved a similarly minimal physical alteration but one which violated the interpretation given to the original target. In contrast, the no-label subjects showed no comparable differentiation between the two kinds of distractors.

EXPERIMENT II

The initial experiment demonstrated the role of semantic comprehension in facilitating free recall of pictures. Having a meaningful name for a picture may facilitate recall because it provides a memorable summary or cue for later free recall. But an interpretation does more than provide a meaningful mnemonic label for a picture; it also causes unification or knitting together of the disparate parts of the picture into a coherent whole or schema.

The second experiment sets out to test more directly the influence of the unifying coherence of an interpretation upon picture memory. The subject was asked to study *pairs* of nonsensical pictures and was later tested by cueing with one member of each pair for recall (in drawing) of the other member of the pair. Again, half of the subjects received no interpretation of the pictures, whereas half heard a phrase which made both pictures and their pairing a meaningful sequence. Examples of picture pairs are shown in the three rows of Figure 2. Their interpretations are (from left to right panels in each pair): (1) rear end of a pig disappearing into a fog bank, and his nose coming out the other side of the fog; (2) piles of dirty clothes, then pouring detergent into the washing machine to wash the clothes; (3) uncooked spaghetti, then cooked spaghetti and meatballs. The hypothesis is that subjects hearing such interpretations during study will show much higher associative recall

than will subjects who study the pictures without the interpretations.

Method

The subjects were 16 university students attending summer school. They were recruited by an advertisement and paid \$1.50 for their participation. They were tested individually, assigned in random alternation to the label and no-label conditions ($n = 8$ per group). The subject was told to learn 30 picture pairs which were shown to him at a rate of one every 12 sec. The pairs were drawn and shown by means of 3 x 5 in. flashcards, one picture on a white card and its mate on a pink card. The subject had been told that in the later recall test he would be shown the picture on the white card and would have to recall (draw) its mate from the pink card. During presentation of each pair, the label subjects heard the experimenter supply an interactive interpretation to bind the picture-pair together. These were descriptions like those given above for the three panels of Figure 2. Following presentation of the 30 pairs, the white deck was shuffled and presented as recall cues at a 20-sec rate. "Gist" sketching of the recalled pictures was emphasized. If the subject had begun his drawing before 20 sec, he was allowed to complete it; otherwise, the next cue was presented after 20 sec. Subjects drew their recall sketches in numbered boxes, nine to a page; they left blank any numbered box for which they could recall nothing to the corresponding cue.

After the cued recall test (conducted without feedback regarding the correctness of subject's recall), the subject received an associative matching test. The 30 white and 30 pink cards were spread out in a random array over the table top. The subject was instructed to scan over the array, looking for the pairs of white and pink pictures he had studied. As pairs were recognized, the two cards were picked up by the subject and handed to the experimenter. The subject continued this pairing until he had selected all pairs he could remember; because they were asked not to guess, many subjects stopped short of pairing off all members. The subject's associative matching score was simply the number of correct pairs he selected from the array before terminating. (The expected correct pairs obtainable by guessing in an associative matching test is about one, regardless of the number of pairs to be guessed at. See Feller, 1957, p. 97.) After completing the matching test, the subject was debriefed and dismissed.

Results

Associative Recall. Again no problem was encountered in scoring correct gist recall of the cued picture. Cued recall averaged 21.75 pictures out of 30 (73%) for the label subjects ($SEM = 1.78$) and 13.13 (44%) for the no-label subjects ($SEM = 2.41$). These percentages differ reliably [$t(14) = 2.87, p < .02$]. The effect is remarkably consistent over items, too: For 22 items the label subjects recalled more than the no-label subjects, for three items the no-label subjects recalled more, and there were five ties. The 22/25 predominance of items with more recalls by label subjects exceeds chance of 50% ($z = 6.55, p < .01$). Thus, the label group is uniformly superior in recall to the no-label group.

Associative Matching. The number of correct matches (of pairs) averaged 27.50 for the label subjects ($SEM = .98$) compared to 16.63 for the no-label subjects ($SEM = 2.80$). These differ reliably [$t(14) = 3.66, p < .01$]. Moreover, the relative gain in recognition

performance above what could be recalled was much larger for the label subjects (70%) than for the no-label subjects (21%). The data show that the label subjects still exhibit superior associative coherence even when all the pictures are available and do not need to be recalled.

DISCUSSION

It has been argued that memory for a picture depends upon the subject achieving a conceptual interpretation of the picture as he views it. The hypothesis is the pictorial analog of that relating sentence recall to comprehension (e.g., Bransford & McCarrell, in press). The point is intuitively obvious once it has been noticed (as are many other "facts" of psychology), and the experiments above are primarily demonstrational in nature. Subjects provided with meaningful interpretations of single doodles show superior free recall. Subjects who hear an interpretation identifying and relating two pictures together show greater coherence of the pictures on later association tests. Although control subjects were probably trying to come up with some sensible interpretation of the pictures, the difficulty of the task precluded much success. Presumably, if we had collected control subjects' attempts at interpretations (recording their "thinking aloud"), those pictures for which they achieved a meaningful interpretation would have been more likely to be recalled (see, e.g., Montague, 1972). The likelihood of this being the case remains to be checked.

One might question whether the associative coherence found in Experiment II is a result merely of identifying the objects in each picture or whether it depends in addition upon providing the meaningful linking relationship between the two pictures of a pair. For some pairs the linking relation was that the pictures denoted different parts of the same object (the pig in Figure 2), different states of an object as it underwent changes (the spaghetti), or different objects associated with a common process (the clothes and washer in Figure 2). We feel these relations are very important for promoting associative coherence of the elements of the pair. To illustrate this point, four further subjects from the same source were tested under the same procedure as Experiment II, except that new pairs were constructed by re-pairing the old pictures in a random manner. As the pair was shown, each picture was separately interpreted (e.g., a pig's tail and detergent pouring into a washing machine). No linking relation other than contiguity was stated for connecting the two contents. These four subjects averaged only 7.75 correct in cued recall ($SEM = 1.89$) and 8.25 correct in associative matching ($SEM = 1.11$). If anything, the scores are lower than those for the controls who studied the original pairs without hearing the objects or relation identified. Quite possibly, this "mispairing" list was so difficult because

semantically related objects appeared in different pairs, creating intrapair interference. Teasing out the several contributors to this poor learning would be a task for further experimentation. The significant fact we wish to glean from this poor recall of mispaired items is that associative coherence depends heavily upon relating the two identified pictures and relatively little upon identifications per se which do not call to mind a known relationship between the two pictures.

How are our results to be related to previous work on picture memory? Previous work on learning of nonsense figures typically used recognition rather than reproduction measures and have been largely concerned with testing hypotheses of acquired distinctiveness or acquired equivalence of forms induced by learning different or the same arbitrary labels for the forms. Of more direct relevance to our results are those by Ellis (reviewed by Ellis, 1973), who found that the learning of "representative labels" to complex polygons enhanced their later recognition. Of course, the "representative labels" were simply a plausible name or interpretation of the figures. The fact that pairing with a representative label makes the pictures more memorable seems quite consistent with our hypothesis relating picture comprehension to memory. Since "association value" or "codability" has been a common variable in research on pictorial memory, one may ask whether our notion of a "semantic interpretation" of a picture is just a fancy name for an association to it. We think not. We intend "semantic interpretation" to be much more specific than the concept of "picture association" suggests. Associations may occur to many surface features of a picture or to fragments of it, all without improving memory for it. Presumably, picture memory would improve with greater "depth of processing," as

does memory for words (Craik, 1973) or faces (Bower & Karlin, 1974). But this implies comprehending the picture, figuring out what conceptualization it expresses or what object it denotes: It means getting the "message" behind the "medium."

REFERENCES

- Bransford, J. D., & Johnson, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, 1972, 11, 717-726.
- Bransford, J. D., & McCarrell, N. S. In D. Palermo and W. Weimer (Eds.), *Cognition and the symbolic processes*. Washington, D.C: Winston, in press.
- Bower, G. H., & Karlin, M. B. Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 1974, 103, 751-757.
- 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.
- Craik, F. I. M. A "levels of analysis" view of memory. In P. Pliner, L. Krames, and T. Alloway (Eds.), *Communication and affect: Language and thought*. New York: Academic Press, 1973.
- Doll, T. J., & Lapinski, R. H. Context effects in speeded comprehension and recall of sentences. *Bulletin of the Psychonomic Society*, 1974, 3, 342-345.
- Ellis, H. C. Stimulus encoding processes in human learning and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*. Vol. 7. New York: Academic Press, 1973.
- Feller, W. *An introduction to probability theory and its applications*. Vol. 1. 2nd ed. New York: Wiley, 1957.
- Gombrich, E. *Art and illusion*. New York: Pantheon, 1960.
- Goodman, N. *Languages of art*. 2nd ed. Indianapolis, Ind: Bobbs-Merrill, 1968.
- Minsky, M. Frame systems. Unpublished manuscript. M.I.T. AI Project, 1974.
- Montague, W. E. Elaborative strategies in verbal learning and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: In research and theory*. Vol. 6. New York: Academic Press, 1972. Pp. 225-302.
- Price, R. *Doodles*. Los Angeles: Price/Stern/Sloan, 1972.
- Riley, D. A. Memory for form. In L. Postman (Ed.), *Psychology in the making*. New York: Knopf, 1963.

(Received for publication June 24, 1974;
revision received July 25, 1974.)