Cross-modal recognition of pictures and descriptions without test-appropriate encoding

COLIN M. MacLEOD
University of Toronto, Scarborough Campus, Scarborough, Ontario, Canada

Droodles are simple line drawings that take on distinct and humorous meanings when coupled with appropriate descriptive phrases. Two experiments examined recognition memory for one element (drawing or phrase) when only the other had been studied earlier. In this way, the interpretive link between drawing and phrase could not be made at the time of encoding. Experiment 1 showed that people could recognize phrases when only the corresponding pictures had been studied, and Experiment 2 showed that people could recognize pictures when only the corresponding phrases had been studied. Thus, cross-modal recognition is successful when test-appropriate encodings could not have been prepared during study. Also, performance was unrelated to imagery ability in both experiments. Apparently, pictures and phrases are encoded and stored so as to permit quite dramatic reinterpretation and comparison during retrieval.

When Carmichael, Hogan, and Walter (1932) published their study on the reproduction of simple pictures, their aim was to demonstrate the reconstructive nature of recall. A very different experience must be remembered when two circles joined by a bar are thought of during encoding as a barbell rather than as eyeglasses. They argued that the descriptive phrase applied to the picture during study altered the memory representation of the picture. What was recalled later was reconstructed from both the pictorial and the verbal inputs. In essence, an interpretation of the picture was remembered.

Less frequently discussed is a follow-up study done by Hanawalt and Demarest (1939). Here, the same reconstructive outcome was observed, but with the verbal description present only at the time of the reproductive recall test. Hanawalt and Demarest (1939, p. 173) concluded that, "If the reproduction can be partly transformed by verbal suggestion in the learning period and by the same means in the recall period, it is very likely that any relevant experience occurring between these two periods or previous to them can be effective *during the construction in recall.*" Viewed more as bias at retrieval than as integration during encoding, this conclusion may seem less provocative than that offered by Carmichael et al. (1932). Perhaps this is why the Hanawalt and Demarest explanation is encountered less often today.

In the past fifteen years, there has been renewed interest in reconstructive memory. One reason is the resurgence of the idea that information is integrated from different sources, an idea prevalent in studies of eyewitness memory (e.g., Loftus, Miller, & Burns, 1978). Another reason is the increased reliance on schema-based views, including the notion of interpreting one stimulus dimension on the basis of another (e.g., Bransford & Johnson, 1972). To test these hypotheses, the cross-modal framework of Carmichael et al. is an obvious choice.

Two types of special materials developed since 1932 are particularly well suited to studying how verbal information influences memory for pictures. Mooney figures (Mooney, 1957) are ambiguous, and can be seen either as faces or as meaningless patterns. As might be expected, such figures are recognized much better on a second viewing when interpreted as faces during the first viewing (Freedman & Haber, 1974; Wiseman & Neisser, 1974). Emphasizing the importance of interpreting the visual stimulus at the time of initial encoding, Wiseman and Neisser (1974, p. 675) wrote, "This result indicates that the nature of the internal representation formed by the subject is a major determinant of recognition accuracy."

Droodles (Price, 1972, 1973) are simple line drawings that take on a uniquely humorous meaning when coupled with appropriate descriptive phrases. Three studies have examined memory for Droodles. In the first, by Bower, Karlin, and Dueck (1975), Droodles were presented for study with or without their meaningful phrases. Later, free recall was better for subjects who had phrases at study; they also made more false alarms to prototype-like distractor pictures on a recognition test. Both the interpretive and reconstructive elements of the Carmichael et al. account were evident in Bower et al.'s explanation.

Klatzky and Rafnel (1976) were concerned that the Bower et al. (1975) result might have been due not to a better interpretation but to the additional retrieval cue provided by the phrase. Consequently, Klatzky and Rafnel's subjects saw the Droodle drawings either without phrases or with one of two types of phrase—meaningful

(the Bower et al. labels) or not meaningful. Although both types of phrase helped cued recall, only meaningful phrases improved free recall. Klatzky and Rafnel argued that meaningful phrases affected interpretation of the drawings, resulting in better encoding, while nonmeaningful phrases merely provided ad hoc associative cues.

Rafnel and Klatzky (1978) developed two kinds of recognition distractors: conceptually similar to the target and conceptually dissimilar from the target. During study, one group was given the Bower et al. meaningful phrases; the other group received labels related to physical details of the drawings. Rafnel and Klatzky observed that the group having meaningful phrases at study performed well in detecting semantic changes on the test, but seemed no better at detecting physical changes than was the group given physical labels. They concluded that semantic coding improved performance only where semantic information was invoked. What is remembered, they argued, is the interpretation of the picture, not the picture itself.

Interpretation is the thread that ties together all of the studies discussed thus far, with the notable exception of Hanawalt and Demarest's (1939) study. It remains quite possible that the effect observed in these newer studies, so like that reported by Carmichael et al., may occur not at encoding but at retrieval. We simply cannot tell without appropriate experiments that manipulate the study-test relation.

Figure 1. Six of the 20 Droodle drawings used in the experiments.

EXPERIMENT 1

Existing arguments suggest that cross-modal recognition should be difficult when formation of a meaningful link is prevented. As Klatzky and Rafnel (1976, p. 719) maintain, "when a meaningful label is given at the time of test only, it fails to facilitate recall." Thus, the major concern of Experiment 1 was how well people retain the meaningless pictures themselves. Subjects saw only the Droodle pictures at study; no verbal descriptions were provided. Figure 1 displays a portion of the set of drawings a subject might see.

The test was an entirely verbal YES/NO recognition test, wherein subjects had to select the descriptive phrases that corresponded to the Droodle pictures seen during study. To discourage initial verbal recoding, subjects were not told of the nature of the recognition test at study. Furthermore, subjects were selected who were unfamiliar with Droodles. To ascertain what role, if any, imagery ability played in remembering the pictures, a test of imagery was administered to all subjects. The overall aim was to determine whether subjects could draw on their memories of the studied pictures when presented with only verbal descriptions at test.

Method

Subjects. Ninety-one University of Washington undergraduates received credit toward their introductory course grades. Based on a postexperiment questionnaire, all were naive concerning Droodles.

Materials. Twenty Droodles and their corresponding descriptive phrases were taken from the two books by Price (1972, 1973). Particular items were chosen on the basis that 21 naive subjects in a pilot study
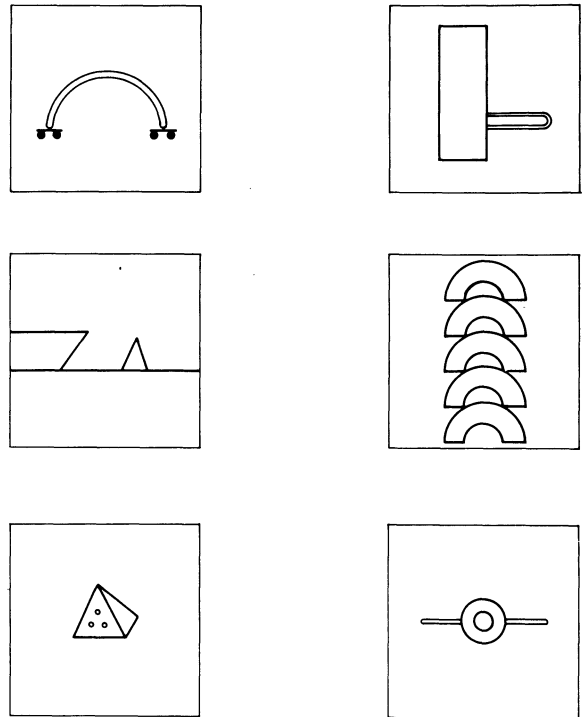
virtually never generated the appropriate descriptive labels for them. All pictures were prepared as black-on-white line drawings and mounted in 35-mm slides for display on a screen 4 ft from the subject.

Procedure. Subjects were tested individually. Every subject saw 10 pictures during study, each picture for 10 sec. The subject was encouraged to remember the pictures for a later test, but the nature of the test was not mentioned. After study, the subject was given a test sheet listing 20 descriptive phrases and asked to indicate which ones matched the 10 pictures studied. No time limit was specified. To obtain a sense for the subject's task, try to recognize which six phrases in Table 1 match the six drawings in Figure 1, without reinspecting the Droodles.

Once this recognition test was completed with no omissions, the subject was asked to provide a confidence rating for each judgment. The scale was as follows: (1) certain, (2) quite sure, (3) somewhat sure, (4) quite unsure, and (5) guessing. The subject was encouraged to use the entire scale, taking as much time as needed. Finally, the subject was asked to review the phrases once more, indicating any which he or she had thought of while studying the corresponding picture.

There were eight subgroups of subjects determined by factorial combination of three binary variables: picture set studied (Set A or Set B), order of pictures at study (two randomizations), and order of phrases on the test (two randomizations). This procedure ensured counterbalanc-

Table 1
Twelve of the 20 Droodle Phrases Used in the Experiments

1. Cannonball sandwich
2. Worm crawling over a razor blade
3. High-rise igloos
4. Midget playing a trombone in a telephone booth
5. Pig coming out of the fog
6. Ship approaching a drowning witch
7. Clam smoking in bed
8. Mexican riding a bicycle
9. Bald men watching a burlesque show
10. Egyptian bowling ball
11. Snake on roller skates
12. Man with his tie caught in an elevator

ing of sets of items and minimized the role of item sequence at study and at test. There were at least 10 subjects in each subgroup.

The last stage was administration of the Vividness of Visual Imagery Questionnaire (VVIQ). The test involves a series of questions about situations in which visual images must be formed and manipulated, and has been shown to have predictive power in other settings (e.g., Marks, 1973). Possible scores range from 32 to 160, from very vivid imagery in all situations to virtually no imagery in any situation.

Results and Discussion

Taking into account all 20 items on the YES/NO recognition test for all 91 subjects, mean d' was 1.12. This was significantly greater than $d' = 0$ [$t(90) = 11.39$, $p < .001$]. On average, subjects thought of appropriate phrases for 14.61% of the pictures at study. Removing these items from the d' calculations, mean d' decreases to 0.92. Nevertheless, this remained significantly different from $d' = 0$ [$t(90) = 9.30$, $p < .001$].

Confidence ratings demonstrated that subjects were quite certain for hits (mean rating of 1.70), but less confident for correct rejections (2.97), misses (3.12), and false alarms (2.82). Discarding the data of the 6 subjects who made no false alarms, analysis of variance confirmed this pattern [$F(3,252) = 64.06, p < .001$]. Only hits were associated with higher confidence; a picture-phrase match was a quite compelling event.

The other relevant data are the VVIQ scores and their relation to recognition performance. VVIQ scores covered virtually the entire range, from 37 to 147, with a mean of 64.4. Recognition scores also showed a wide range, from a d' of -1.02 to a d' of 3.16. Thus, there was no restriction of range, and the sample size was large enough for a fair test. Yet the correlation was a nonsignificant $r=.06$. If the VVIQ provides a reasonable measure of imagery ability, then that ability is not an important determinant of performance in this study.

Thus, subjects were able to recognize a description of a picture seen earlier even when that description had not been available to encode while studying the picture. In all likelihood, a completely unrelated verbal description would have been encoded at study. Apparently, subjects had some sort of visual information available in memory for at least some of the pictures, and were able to use the phrases to retrieve this stored visual information and to make the necessary comparison. In line with this account, subjects reported that they felt they could not do the test initially, but that once they caught on, they were able to inspect their memories for the pictures and then compare the remembered pictures to the phrases.

EXPERIMENT 2

The design of Experiment 2 was basically the opposite of that of Experiment 1. Could subjects use pictures at the time of test to probe their memories for studied descriptive phrases? Subjects saw only the phrases at study. The test, however, was a YES/NO recognition test of Droodle pictures only, with subjects required to choose those pictures that matched the phrases studied. As be-

fore, to discourage recoding during study, subjects were told nothing of the form of the upcoming test.

Again, the VVIQ was administered. Also, a postexperiment questionnaire was included to check on subjects' familiarity with Droodles and their likelihood of recoding the studied phrases into test-appropriate pictures. The overall aim was to examine whether pictures and phrases could be compared in memory when this comparison was not anticipated and appropriate recoding was not possible during study. If so, this would confirm and extend the result of Experiment 1.

Method

Subjects. Eighty University of Washington undergraduates took part to obtain credit toward their introductory course grades. Based on the postexperiment questionnaire, one subject's data were excluded.

Materials and Design. The materials were identical to those in Experiment 1, except that the phrases now were typed and prepared as 35-mm slides for study. The design was the same 2 × 2 × 2 used in Experiment 1 with 10 subjects per counterbalancing subgroup.

Procedure. Subjects were tested in small groups. They saw 10 phrases during study, each for 10 sec, and were instructed to remember the phrases for later recall. The picture recognition test was not mentioned. After study, both tests were administered. In free recall, subjects were asked to recall the wording of the phrases as accurately as possible. They were allowed as long as they wanted.

The second test was a 20-picture YES/NO recognition test, with the pictures corresponding to the 10 studied phrases mixed together with the 10 pictures corresponding to unstudied phrases. Each picture was exposed for 30 sec, during which time subjects performed four tasks: (1) deciding whether the picture matched a studied phrase, (2) making a confidence judgment on the same scale used in Experiment 1, (3) trying to write down the remembered phrases for all YES responses, and (4) indicating whether any picture had been imagined at the time of study. No omissions were permitted on the first two parts. All subjects completed the VVIQ test of imagery following the session.

Results and Discussion

Calculated over all 20 test pictures in YES/NO recognition for all 79 subjects, mean d' was 1.58. This value was significantly greater than zero [$t(78) = 13.77, p < .001$], indicating that subjects succeeded in identifying pictures at test that matched phrases studied earlier. On average, subjects thought of appropriate pictures for 15.44% of the studied phrases, a rate comparable to that of Experiment 1. Removing these items from the calculations, mean d' becomes 1.34, still significantly greater than zero [$t(78) = 11.30, p < .001$].

Confidence ratings also followed the same pattern as Experiment 1. Subjects were quite certain of their hits (mean rating of 1.72), but considerably less sure for correct rejections (3.04), misses (3.22), and false alarms (3.15). Discarding the data of the 22 subjects missing an entry (either a miss or a false alarm), analysis of variance confirmed the pattern [$F(3,168) = 41.38, p < .001$]. A phrase-picture match was very convincing when recognized. That subjects provided the correct phrase for 86.7% of their hits reinforces this claim.

Once again, the relation between imagery ability and memory performance was investigated. The recall test was included primarily for this purpose. Subjects recalled 3 to 9 of the 10 phrases (mean of 6.96), and VVIQ scores

ranged from 37 to 138 (mean of 66.5). Given the large sample size, there was ample room for a correlation to emerge. However, the r of .11 was nonsignificant and in the wrong direction. Nor was VVIQ correlated with d' in recognition. Despite a good range—from a d' of $-.52$ to a d' of 3.16—the correlation was $r = .04$. Imagery ability is not an important determinant of recalling these phrases or of comparing the pictures with the phrases.

Consistent with Experiment 1, subjects were able to recognize stimuli cross-modally, this time selecting pictures correctly after studying phrases originally. They did this despite being unable to generate test-appropriate encodings during study, and having only one picture in view at a time during test. Some kind of translation of one form into the other (or both into a common third form) must occur at the time of test, as was suggested by subjects' comments on the questionnaire.

To summarize, although they could not anticipate the form of the recognition test, and hence could not prepare test-appropriate encodings during study, subjects could recognize a verbal description expressed as a line drawing. The suggestion is that initial encoding was in a form flexible enough to permit successful cross-modal comparison at the time of test. Experiment 2 generalizes this conclusion so that it applies in both directions with respect to pictures and their verbal descriptions.

GENERAL DISCUSSION

The predominant explanation of better memory for Droodle drawings given meaningful labels places emphasis at the encoding stage. Following Carmichael et al. (1932), the argument is that a meaningful interpretation leads to a superior encoding of an otherwise meaningless form. However, lest this be thought of as a more accurate encoding, studies have shown that the remembered information is actually truer to the meaning than to the form (e.g., Bower et al., 1975; Carmichael et al., 1932; Rafnel & Klatzky, 1978). Once a stimulus is interpreted, that interpretation, not the actual form, is what is remembered.

Because virtually all of the existing studies have manipulated meaningfulness only at study, it should come as no surprise that accounts emphasize encoding. Yet Hanawalt and Demarest (1939) provide a real alternative. Perhaps, they reason, the meaningful information exerts its effect regardless of whether it is presented during study, during the retention interval, or at the time of test. Their results confirm this conjecture, and point to retrieval as the critical stage. When trying to recover a meaningless form, information about how to think about that form will be influential. However, the interpretation need not be encoded with the meaningless stimulus at study.

The present experiments provide support for Hanawalt and Demarest's ideas and also take them a step further. People can remember half of an interpretable couplet and connect it to the other half later, even when they did not expect to have to do this. Whether remembering disconnected phrases or meaningless line drawings, subjects in these experi-

ments successfully recognized one on the basis of the stored representation of the other. Encodings were general enough to permit a wholly new interpretation to be recognized later.

What role does imagery play here? If images mediate the drawing-phrase comparison, it seems reasonable to expect a relationship between imagery ability and recognition performance. No such relationship was observed. Of course, it might be argued that all subjects had sufficient skill for the task, but this seems unlikely given the range of VVIQ scores. Results like these highlight the difficulties involved in aligning stimulus differences, instruction differences, and subject differences in imagery research (Slee, 1978).

Taken together, the two experiments reported here are consistent with three main conclusions. First, stimulus encoding is certainly flexible enough in some situations to allow unexpected cross-modal comparisons to be carried out successfully in memory. At the same time, the complexity of the retrieval operations must be recognized. Second, in this particular drawing-phrase comparison, imagery ability is not an important determinant of recognition performance. And third, it is conceivable that storage of visual forms may preserve their visual characteristics when interpretation is not required during encoding.

REFERENCES

BOWER, G. H., KARLIN, M. B., & DUECK, A. (1975). Comprehension and memory for pictures. *Memory & Cognition*, **3**, 216-220.

BRANSFORD, J. D., & JOHNSON, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, **11**, 717-726.

CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, **15**, 73-86.

FREEDMAN, J., & HABER, R. N. (1974). One reason why we rarely forget a face. *Bulletin of the Psychonomic Society*, **3**, 107-109.

HANAWALT, N. G., & DEMAREST, I. H. (1939). The effect of verbal suggestion in the recall period upon the reproduction of visually perceived forms. *Journal of Experimental Psychology*, **25**, 159-174.

KLATZKY, R. L., & RAFNEL, K. (1976). Labeling effects on memory for nonsense pictures. *Memory & Cognition*. **4**, 717-720.

LOFTUS, E. F., MILLER, D. G., & BURNS, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 19-31.

MARKS, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, **64**, 17-24.

MOONEY, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology*, **11**, 219-226.

PRICE, R. (1972). *Droodles*. Los Angeles: Price/Stern/Sloan.

PRICE, R. (1973). *Droodles No. 2*. Los Angeles: Price/Stern/Sloan.

RAFNEL, K. J., & KLATZKY, R. L. (1978). Meaningful-interpretation effects on codes of nonsense pictures. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 631-646.

SLEE, J. (1978). The consistency of different manipulations of visual imagery: A methodological study. *Australian Journal of Psychology*, **30**, 7-20.

WISEMAN, S., & NEISSER, U. (1974). Perceptual organization as a determinant of visual recognition memory. *American Journal of Psychology*, **87**, 675-681.