

The effect of repeated tests on recognition memory for pictures and words

JOAN GAY SNODGRASS and PHYLLIS McCLURE BURNS
New York University, New York, New York 10003

To determine whether the advantage of picture memory over word memory would be maintained over repeated tests with the same material, eight subjects participated in six sessions of a recognition memory experiment. In each session after the first, half of the studied items and half of the lures from the previous session were switched to the opposite category. Recognition memory, whether measured by hits, false alarms, A' , or confidence ratings, remained at a constant high level across the six sessions, and pictures maintained their advantage over words. There was a small sequential advantage of maintaining an item in the same category from one session to the next, but this was only reliable for new items. We concluded that the advantage of pictures over words in memory is not because of their novelty, and that situational familiarity of the type manipulated here does not appear to operate analogously to naturally occurring familiarity.

Although memory for the pictured form of a concept has typically been found superior to memory for that concept's name, the locus of the picture-superiority effect has remained elusive. One possibility, particularly relevant to the recognition memory task, is that pictures are relatively more unique in the life experience of subjects than words. It is a well established finding that rare words are recognized better, but recalled more poorly, than frequent words, when familiarity is measured by counts of words in print. This paradoxical effect of word frequency on memory has received a fair amount of attention in the literature because of its implications for the relationship between recognition and recall (Glanzer & Bowles, 1976; Kintsch, 1970). Pictures do not show such a paradoxical effect: Typically, pictures are both recalled and recognized better than words (Paivio, 1969, 1971). Nonetheless, it seems apparent that the typical set of pictures used in recognition memory studies will be more unfamiliar to subjects than a typical set of words, if for no other reason than the fact that there are more different ways to represent a picture of a concept than to represent its name.

Within a strength theory model of recognition memory of the sort developed by Norman and Wickelgren (1969), recognition performance can be good if old items are high on a strength or familiarity continuum, new items are low on the familiarity continuum, or both. In their analysis of recognition memory for frequent and infrequent words, Glanzer and Bowles (1976) found that both states held for rare as compared with common words: Old rare words

This research was supported by National Institute of Mental Health Research Grant MH-20723 to the first author. Requests for reprints should be sent to Joan Gay Snodgrass, Department of Psychology, New York University, 6 Washington Place, New York, New York 10003.

were higher on the familiarity continuum than old common words, and new rare words were lower on familiarity than new common words.

A similar question can be asked about comparisons of pictures and words. Pictures can derive their advantage by being more familiar as old items or more unfamiliar as new items than words. One way to test whether pictures derive their advantage by their unfamiliarity as new items is to draw new items from pools of items to which subjects have previously been exposed. If pictures derive their advantage from the uniqueness of new items, repeated tests with the same items should produce a larger decrement in recognition memory for pictures than for words.

METHOD

Subjects

Eight females associated with New York University, two secretaries, one graduate student, and five undergraduates, were paid for their participation in the study. They ranged in age from 18 to 28 years.

Apparatus and Materials

A set of 80 slides, 40 pictures and 40 words, were used as stimuli. The pictures were simple line drawings of common concepts (simple scenes, animals, people, and objects) and the words were the names of the pictures. The 40 pictures referred to different concepts from the 40 words, and the two sets were matched as closely as possible on category membership, apparent familiarity, and complexity of the concept's pictured form.

A Kodak carousel projector was used to project the stimuli, and an L-shaped geometric puzzle and a set of tangrams served as distraction tasks.

Procedure

Subjects were run in groups of from three to eight. Each subject participated in six sessions, separated by between 1 and 5 days, depending upon availability of the subjects. However, all subjects completed the six sessions within a 3-week period.

At the beginning of each session, subjects were shown a study set consisting of 20 pictures and 20 words for a total of 40 study items. The pictures and words were presented in a strictly alternating sequence via a slide projector, and each stimulus was exposed for 5 sec. After the study phase, subjects were given a 5-min break, during which they attempted to solve one of six geometric puzzles. For the first session, the geometric puzzle consisted of an L-shaped figure that had to be divided into four equal parts. For succeeding sessions, subjects had to form a designated geometric figure from a set of seven pieces (tangrams).

The test phase consisted of presenting the 40 old items randomly interspersed with a set of 40 new items. The old items were presented in a different random order from their presentation during the study sequence, and for the test sequence, pictures and words were randomly, rather than alternately, interspersed.

Since the total population of picture and word stimuli consisted of 80 items, by the end of the first session all subjects had seen all of the items that would be presented in the subsequent five sessions, either as old items or as new items. For each successive session, half of the study items from the previous session remained study items and half became lures, while half of the lures became study items and half remained lures, and the distribution of changed-state and same-state items was equal across pictures and words. The sequence of study-test sessions was designed so that each picture and each word appeared exactly three times as a study item and three times as a lure.

During the test, each item was presented for about 10 sec. Subjects recorded on data sheets whether they thought the item was old or new, and then rated the confidence of their judgment on a 3-point rating scale. The combination of the old-new judgment and the 3-point rating scale was converted into a 6-point rating scale for the purpose of data analysis, in which 1 signified a very confident "old" judgment, and 6 a very confident "new" judgment.

After the first session, subjects were informed that in successive sessions, they would be shown exactly the same set of pictures and words, but that the items designated as old and new would change from session to session.

RESULTS

The hit and false alarm rates for old pictures and words for each subject were converted into the nonparametric measure A' . A' is the average area under a variety of MOC curves and is relatively neutral with respect to the underlying model proposed for recognition memory. It further has the advantage of being defined when performance is perfect (i.e., when the hit rate is 1.0 and the false alarm rate is 0, A' equals 1.0).¹

Table 1 presents average A' values for pictures and words as a function of sessions. While picture memory performance appeared to be better than word memory

Table 1
Mean A' Values for Picture and Word Recognition Memory by Session

| | Session | | | | | |
|----------|---------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Pictures | .969 | .983 | .980 | .974 | .984 | .970 |
| Words | .922 | .930 | .966 | .960 | .950 | .947 |

Table 2
Mean Confidence Ratings for Old and New Pictures and Words by Sessions

| | Session | | | | | |
|--------------|---------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Old Pictures | 1.37 | 1.16 | 1.27 | 1.37 | 1.26 | 1.29 |
| New Pictures | 5.60 | 5.57 | 5.52 | 5.47 | 5.33 | 5.33 |
| Old Words | 1.71 | 1.52 | 1.48 | 1.59 | 1.57 | 1.51 |
| New Words | 5.29 | 5.08 | 5.39 | 5.27 | 5.08 | 5.12 |

performance, there was no apparent decrement in performance over sessions for either pictures or words. In particular, there was no evidence that the cumulative effects of familiarity (as measured by progress over sessions) was any different for pictures than for words. A two-way analysis of variance on the arc-sine transformations of the A' values showed that the effect of pictures vs. words was significant [$F(1,7) = 12.31$, $p < .01$], while neither the effect of sessions nor the interaction of sessions with stimulus type was significant (both F s less than 1.0). Exactly the same results were obtained for the arc-sine transformations of the hit values, and the arc-sine transformations of the false alarm rates. For hits, the F for pictures vs. words was 13.53 ($p < .01$), and for false alarms, the F was 6.23 ($p < .05$). For neither hits nor false alarms was either sessions or the interaction significant.

Confidence Ratings

Table 2 presents the average confidence ratings for old and new pictures and words as a function of sessions. The maximum confidence value for old items is 1.0, and the maximum confidence value for new items is 6.0. Of course, the maximum values can only be attained by subjects with perfect performance as well as maximum confidence on each judgment. Again confidence did not seem to be affected by familiarity, although confidence ratings were higher for pictures than for words. In addition, confidence did not seem to be different for old vs. new items.

The results of a three-way repeated measures analysis of variance, with one factor pictures vs. words, a second factor old vs. new, and a third factor sessions, confirmed these impressions. Before performing the analysis of variance, the new item confidence ratings were reflected (by subtracting each rating from 7) in order to convert both rating scales to the same maximum confidence of 1.00. The only main effect of significance was pictures vs. words [$F(1,7) = 25.17$, $p < .001$]. None of the other main effects and none of the interactions reached significance. Thus, we come to the same conclusions from the analysis of confidence ratings that we did from the analysis of overall performance.

Sequential Effects

Although no overall effect of familiarity across the six sessions was observed, it seemed possible that there

Table 3
Probabilities of Correct Recognition for Single, Double, and Triple Sequences of Sessions

| | Old Items | | | | | | New Items | | | | | | | |
|----------|-----------|------|------|------|------|------|-----------|------|------|------|------|------|------|------|
| | O | OO | ON | OOO | OON | ONO | ONN | N | NN | NO | NNN | NNO | NON | NOO |
| Pictures | .958 | .967 | .959 | .969 | .969 | .958 | .953 | .954 | .974 | .934 | .956 | .976 | .930 | .949 |
| Words | .898 | .913 | .901 | .909 | .895 | .900 | .906 | .921 | .944 | .899 | .956 | .965 | .932 | .886 |

Note—The first letter of each sequence is the state of the item in the target session, the second letter is its state in the preceding session, and the third letter is its state in the next preceding session.

might be some transient local effects. That is, an old item that appeared twice in a row as a study item, or a new item that appeared twice in a row as a lure, might show some advantage over an old item that appeared in the previous session as a lure, or a lure that appeared in the previous session as an old item. Because after the first session, subjects became desynchronized by session number, the particular items repeated as old and new items varied across subjects, so that any sequential effects observed cannot be attributed to item differences.

Table 3 presents probabilities of correct item recognition for single sessions and for successive pairs and triplets of sessions. The first letter of each sequence represents the state of the item on any particular session, and the following letters represent the state of that item on the previous or two previous sessions. For example, the probability listed for the sequence ONO represents the conditional probability of a hit given that the studied item appeared as a lure in the previous session and as a study item in the session before that one.

The unconditional probabilities O and N represent the hit and correct rejection probabilities averaged across all six sessions. Since the unconditional probabilities include first session items, the conditional probabilities for pairs are subsets of the unconditional probabilities, and the conditional probabilities for triplets are subsets of the pair conditional probabilities.

Table 3 reveals some small effects of items repeated in the same state; the effects were larger for pairs than for triples of sessions, and the effect was more marked for new items (i.e., for items presented twice in a row as lures) than for old items (items presented twice in a row as study items). To test whether the sequential effects were significant for pairs of sessions, a three-way analysis of variance with repeated measures was performed on the pairs data. The only significant main effect was the difference between pictures and words [$F(1,7) = 9.50, p < .05$]. There was no main effect of either old vs. new items ($F < 1$) or of repeated items per se [$F(1,7) = 4.50, n.s.$]. However, there were two significant double interactions. The interaction of picture vs. word and old vs. new was significant [$F(1,7) = 5.78, p < .05$], reflecting the fact that old pictures were recognized somewhat better than new pictures, whereas new words were recognized somewhat better than old words. There was also a significant

interaction between whether an item was old or new and whether it was repeated or nonrepeated [$F(1,7) = 9.17, p < .05$], reflecting the fact that repetition of new items produced better performance than repetition of old items. A contrast of the repeated-nonrepeated difference for new items revealed the effect of repeating lures to be highly significant [$F(1,7) = 31.26, p < .001$]. Thus, new items appear to benefit much more from being new on the previous session than old items benefit from being old on the previous session.

DISCUSSION

In contrast to studies that manipulate familiarity of verbal materials by the use of naturally occurring stimuli (i.e., frequency in print), the manipulation of familiarity in the present study had virtually no effect on performance over sessions, whether measured by A', hit rate, false alarm rate, or confidence ratings. A small effect was observed for pairs of successive sessions, but only for the recognition of new items.

Several investigators have suggested that the word-frequency effect on memory may be attributable to aspects peculiar to the words themselves. Glanzer and Bowles (1976) have suggested that the higher association value of high-frequency words, or the fact that familiar words have more meanings, can account for both the recognition advantage of rare words and the recall advantage of frequent words. Landauer and Streeter (1973) have shown that rare words are structurally different from frequent words. Carroll and White (1973) have suggested that the frequency effect in memory is mediated via the effect of age of acquisition, and in and of itself is an ineffective variable when age of acquisition is controlled. Kinsbourne and George (1974) factorially varied both association value and frequency and found that frequency in print had significant and opposite effects on recognition and recall when association was controlled. They also showed that when situational frequency of words was manipulated by preexposing one group of subjects to the set of words prior to the experiment in an incidental learning task, subjects preexposed to words showed poorer recognition than those not so exposed.

The last experiment appears to be one of the few that varied situational frequency in addition to, or instead of, naturally occurring frequency. However, Kinsbourne and George's (1974) manipulation of situational frequency was different from ours, in which frequency was varied across a series of sessions. In the present experiment, subjects were informed that they would see the same items over and over again, and this knowledge may have helped them develop strategies that would be immune to interference from previous sessions. In fact, several subjects spontaneously reported that they attempted to remember the old items by developing an elaborate structure, such as a story or scene, that would serve to associate the old items together. If such a strategy was used, it appears to have been more effective with pictures than with words.

In summary, there is no evidence from the present study that subjects are better able to discriminate old pictures from new pictures because of the pictures' uniqueness.

REFERENCES

- CARROLL, J. B., & WHITE, M. N. Word frequency and age-of-acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, 1973, **25**, 85-95.
- GLANZER, M., & BOWLES, N. Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 21-31.
- KINSBOURNE, M., & GEORGE, J. The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 63-69.
- KINTSCH, W. *Learning, memory, and conceptual processes*. New York: Wiley, 1970.
- LANDAUER, T. K., & STREETER, L. A. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 119-131.
- NORMAN, D. A., & WICKELGREN, W. Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology*, 1969, **6**, 241-208.
- PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241-263.
- PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston, 1971.

NOTE

1. A' is calculated from the hit (H) and false alarm (F) rates as follows: for $H \geq F$, $A' = \frac{1}{2} + [(H - F)(1 + H - F)] / [4H(1 - F)]$; for $H < F$, $A' = \frac{1}{2} - [(F - H)(1 + F - H)] / [4F(1 - H)]$.

(Received for publication January 14, 1978.)