

Evidence that broader processing facilitates delayed retention

WILLIAM F. BATTIG and GILLES O. EINSTEIN
University of Colorado, Boulder, Colorado 80309

After rating words on one to three of either three highly-correlated dimensions (concreteness, imagery, and categorizability) or three relatively uncorrelated dimensions (concreteness, pleasantness, and number of word features), 48 college students were tested without warning for 48-h delayed recall and recognition of the 36 words. The latter broad-processing group showed significantly better recognition than the former narrow-processing group. Group differences in recall, however, were nonsignificant, and the narrow group was slightly superior on words rated for all three dimensions. The results indicate delayed recognition, if not recall, to be better for words which have been processed more broadly or elaborately. Memory is further suggested to be more closely dependent upon breadth or elaboration than depth of semantic processing.

That memory for verbal materials should be improved if these have undergone broader or more elaborate processing is clearly implied by the recent expansion of the levels-of-processing approach to incorporate elaboration as well as depth of processing as critical determinants of memory (e.g., Craik & Tulving, 1975). With the notable exception of Klein and Saltz (1976), however, there is little evidence for such facilitative effects of more elaborate processing.

More specifically, Jenkins and his associates have found no evidence that words processed either by two semantic tasks (Hyde, 1973) or a semantic and nonsemantic task (Walsh & Jenkins, 1973) are recalled any better than words processed with a single semantic task. The semantic-nonsemantic result was also obtained for 24-h delayed recognition by Schwarz (1976). All of these experiments, however, required that both types of processing be performed during a single relatively short (4- to 9-sec) presentation and (except for Schwarz) used very short retention intervals.

The present paper reports an exploratory use of quite different and hopefully more appropriate techniques for investigating the effects of breadth of processing upon delayed retention. The major distinguishing features (as described in more detail below) include (1) multiple presentations of the same word, with a different type of processing during each presentation; (2) within-subject variations from one to three types of processing on a given word; (3) a longer (48-h) retention interval followed by both recall and recognition tests; (4) variations in breadth of processing through the use of highly intercorrelated (concreteness, imagery, and

categorizability) and relatively uncorrelated rating dimensions (concreteness, pleasantness, and number of word features).

After the present data had been collected, Klein and Saltz (1976) reported poorer immediate recall for words processed on two highly correlated than two less correlated semantic dimensions selected from happy-sad, big-little, pleasant-unpleasant, and fast-slow. Their procedure, however, was quite different in requiring subjects to shift across four to six different types of ratings with only 24 words, with two ratings during a single presentation and somewhat variable short retention intervals. Thus, the present experiment represents a significant extension of the paradigm and the results of Klein and Saltz (1976).

METHOD

Each of 48 students in an introductory research methods course provided data during two classroom sessions 48 h apart. During the first session, all students were given three rating sheets each, with different rating instructions and 24 words to be rated accordingly, so that all 24 words on each sheet were rated for the same one of the three rating dimensions. These 72 ratings on 36 different words (which were identical for all subjects) provided that 12 words each appeared on only 1, 2, or all 3 of these sheets and thus were rated for 1, 2, or 3 different dimensions.

During the next class 2 days later, students first were given (without prior warning) recall sheets for the 36 words they had rated. After completing and turning in the recall sheet, each subject was given a sheet containing all 36 rated words intermixed randomly with 36 distractors (each of which was either semantically or structurally similar to at least one of the 36 rated words). For each of these 72 words, subjects were instructed to circle one of four symbols: Y (quite sure you rated the word), YG (unsure, but best guess is that you rated the word), NG (unsure, but best guess is that you did not rate the word), or N (quite sure you did not rate the word). Number of times (one to three) the word had been rated was also estimated for words eliciting Y or YG responses. Both the recall and recognition tests were subject paced, as were the previous word ratings.

This is Publication 68 of the Institute for the Study of Intellectual Behavior, University of Colorado, and was supported by Grant BNS 72-02084 from the National Science Foundation. We thank Lewis O. Harvey, Jr., for making possible the present signal detection (d') analysis of the present data.

The groups differed only in whether their three rating dimensions were highly interrelated or relatively uncorrelated. The correlations were based upon a large word-scaling project currently nearing completion at Colorado, wherein 2,854 words were rated for each of seven dimensions. Three dimensions (concreteness, imagery, and categorizability) all yielded pairwise correlations of at least .88, and therefore were selected as the rating dimensions for the narrow-processing group. Another set of three dimensions (concreteness, pleasantness, and number of features or attributes of the word) were all intercorrelated .39 or less, and thus were selected as rating dimensions for the broad group. Elimination of those subjects who failed to complete all tasks in both sessions left 25 in the broad group and 23 in the narrow group (one of the latter failing to complete the recall task).

Within each group, each rating dimension was used for each of the three sets of 12 words with approximately equal frequency and the order in which these dimensions were rated was also counterbalanced across subjects. The 24 words on each rating sheet always included all 12 words being rated for all 3 dimensions, 8 words rated for 2 dimensions (meaning that 4 of these words were rated for each possible combination of 2 dimensions), and 4 words rated only for that single dimension. Although the same 12-word set was always rated 3 times (as well as twice and once), these 3 sets of 12 words were carefully chosen so that each was closely comparable to the other 2 sets in ratings on all 7 dimensions from the aforementioned word norms (including also meaningfulness and familiarity). Moreover, these words were selected from those used in a previous study (Battig, Note 1), and thus ranged between high and low extremes in both imagery and familiarity. The 12 words rated 3 times were ANGER, BAGPIPE, BEGGAR, BLACKSMITH, BUFFOON, COLD, DEBACLE, HIGH, MERMAID, REASON, TIME, and TRUTH. Words rated twice were BASEMENT, BODY, BUTTERFLY, CHAIR, DOLL, HENCHMAN, IMPROPRIETY, MISCONCEPTION, PREFIX, RIGHT, SOUL, and TEMPLE. Once-rated words were ABASEMENT, AUTHOR, BARD, BIBLE, BREAD, CHANCE, COMMAND, DAWN, LABOR, TOMB, TWEEZER, and WISH.

RESULTS

In order to evaluate generality across words as well as subjects, separate analyses of variance using words and subjects as the random error variate were performed both on proportions of words recalled correctly and rated words correctly recognized (with Y responses). Since the word and subject analyses both yielded generally consistent results, only the latter will be reported except where substantially different findings emerged from the word analysis.

Table 1
Mean Proportion Correct in 48-h Recall and Recognition

	Number of Ratings			Total
	One	Two	Three	
	Recall			
Narrow Group	.030	.121	.292	.148
Broad Group	.050	.167	.263	.160
Total	.041	.145	.277	.154
	Recognition			
Narrow Group	.431	.638	.837	.635
Broad Group	.510	.750	.887	.716
Total	.472	.696	.863	.677

As shown in Table 1, both recall and recognition improved markedly as number of ratings increased from one to three (all $F_s > 79$, $p < .001$). Although the broad group showed slightly better overall recall than the narrow group, Table 1 shows this to be slightly reversed for words rated three times. The overall group difference, therefore, was nonsignificant ($F < 1$), but the interaction of Groups by Number of Ratings approached significance for the subjects analysis [$F(2,90) = 2.44$, $p < .10$] and was significant for the words analysis [$F(2,33) = 3.46$, $p < .05$].

Recognition of previously rated words, however, was consistently and significantly superior for the broad over the narrow group [$F(1,46) = 5.18$, $p < .05$]. Calculations of d' sensitivity measures for each subject, based on frequencies of the four types of recognition responses (see Method) for both rated and distractor words (Dorfman & Alf, 1969), showed an even more impressive superiority of the broad (1.89) over the narrow group (1.45) in recognition performance [$F(1,46) = 7.93$, $p < .01$]. The interaction of Groups by Number of Ratings, however, fell far short of significance [$F(2,92) = 1.69$, $p > .10$].

Frequency judgments of number of previous ratings was relatively poor, being overestimated for correctly recognized words rated once (1.35), with even larger underestimates for words rated twice (1.58) and three times (1.94). Nonetheless, analysis of mean frequency judgments for the 36 rated words showed the differences to be highly significant [$F(2,33) = 14.88$, $p < .001$]. There was no evidence, however, that facilitated recognition by the broad group reflected greater perceived rating frequencies. In fact, when corrected for group differences in numbers of rated words recognized correctly, the narrow group actually gave significantly larger mean frequency judgments (1.67) than the broad group (1.57) [$F(1,33) = 9.21$, $p < .005$], with no interaction of groups with actual rating frequencies ($F < 1$).

DISCUSSION

Although broader processing in the form of multiple ratings on relatively uncorrelated dimensions clearly facilitated delayed recognition, word recall was only slightly facilitated thereby. The latter may reflect merely insensitivity due to the present low level of overall recall performance (15.4%). Alternatively, recognition may be more strongly influenced by variations in breadth or elaboration of processing than is recall, as suggested also by previous failures to find such effects on recall (Hyde, 1973; Walsh & Jenkins, 1973).

Since total processing was equivalent for the present broad and narrow groups, the latter could be viewed as producing greater depth of processing than the broad group. Thus, the present results could be interpreted as indicating breadth or elaboration of semantic processing to be more important, at least for recognition memory, than is depth of processing.

Somewhat problematic, however, is the present finding that any recall superiority for broad over narrow processing was localized primarily in words rated on only two dimensions. Although words rated on all three dimensions should show maxi-

mal benefit from broader processing, these words actually showed slightly poorer recall than under the present narrow-processing conditions. Table 1 also suggests a similar trend for recognition, but this is much less pronounced and clearly non-significant. Moreover, when a possible recognition ceiling effect for words rated three times is taken into account, by comparing the actual broad-group superiority to its maximum possible superiority over the narrow group, this relative improvement is virtually identical for thrice-rated (30.7%) and twice-rated words (30.9%).

Since words rated only twice also differed as to which two of the three rating dimensions were used, the present maximal broad-processing facilitation for such words suggests that ratings on more distinct dimensions may aid in identifying and/or distinguishing between words. Examination of the present words, however, gave no indication that those words showing the greatest recall and/or recognition facilitation for broad over narrow conditions (e.g., DOLL, IMPROPRIETY) are different in discriminability or any other obvious way. Moreover, there was little consistency across recall and recognition tasks either in overall performance level or in direction and magnitude of broad-narrow performance differences, when compared for individual words.

Since the present broad-narrow retention differences varied considerably across recall and recognition tasks as well as words, and changes in these differences were marginally significant only for recall, any further speculation about possible interpretations of such differences appears unwarranted without further research. Nonetheless, increased memory facilitation as number and/or range of different types of processing are increased would appear to follow not only from the Craik-Tulving (1975) elaboration-of-processing view, but also from Saltz's closely related cognitive-space model (Klein & Saltz, 1976) and from feature-sampling models (e.g., Pellegrino & Salzberg, 1975). Extremely broad or divergent types of multiple processing, however, may well produce sufficient encoding discrepancies or conflicts to become less effective for memory than intermediate levels of breadth or elaboration. Thus, research is clearly needed which involves systematic variations in number as well as breadth of multiple types of processing, including also repetitions of the same type of processing.

REFERENCE NOTE

1. Battig, W. F. *Rated numbers of features for words varying in rated frequency and imagery*. Paper presented at the meetings of the Psychonomic Society, November 7, 1975, Denver, Colorado. (Also Tech. Rep. No. 46, Program on Cognitive Factors in Human Learning and Memory, Institute for the Study of Intellectual Behavior, University of Colorado, 1975.)

REFERENCES

- CRAIK, F. I. M., & TULVING, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, 104, 268-294.
- DORFMAN, D. D., & ALF, E., JR. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating method data. *Journal of Mathematical Psychology*, 1969, 6, 487-496.
- HYDE, T. S. Differential effects of effort and type of orienting task on recall and organization of highly associated words. *Journal of Experimental Psychology*, 1973, 97, 111-113.
- KLEIN, K., & SALTZ, E. Specifying the mechanisms in a levels-of-processing approach to memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 671-679.
- PELLEGRINO, J. W., & SALZBERG, P. M. Encoding specificity in associative processing tasks. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 1, 538-548.
- SCHWARZ, W. *Encoding and retrieval processes in recognition memory: The role of input processing, output processing, and retention interval for overall recognition, false recognition of related distractor words, and recognition context effects*. Unpublished PhD dissertation, University of Colorado, 1976.
- WALSH, D. A., & JENKINS, J. J. Effects of orienting task on free recall in incidental learning: "Difficulty," "effort," and "process" explanations. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 481-488.

(Received for publication February 21, 1977.)