

The generation effect: A reflection of cognitive effort?

PAULA T. HERTEL

Trinity University, San Antonio, Texas

In incidental learning tasks, subjects *generated* words from anagrams or incomplete sentences, *verified* that the words solved the anagrams or fit in the sentences, or *evaluated* which rule had been used to construct the word from the anagram or sentence. Latencies in responding to a tone during these trials were used as a measure of cognitive effort. The results indicated that, in comparison to verification, the relatively effortless generation of words benefited memory, but the effortful decisions about the rules did not. Clearly, cognitive effort does not always announce better memory.

In the early 1980s, the literature on memory processes had begun to promise that thinking hard would pay off in better memory. Materials encountered in more difficult tasks are remembered better than those in easier tasks (Eysenck & Eysenck, 1979; Griffith, 1976; Jacoby, Craik, & Begg, 1979; Johnston & Uhl, 1976; Krinsky & Nelson, 1981; Tyler, Hertel, McCallum, & Ellis, 1979; Walker, Jones, & Mar, 1983). Although the promise occasionally was broken by counterexamples of this positive relationship between cognitive effort and memory (see Britton, Westbrook, & Holdredge, 1978; Kellogg, 1984), it was tempting to view cognitive effort as a predictor of recall, as a consequence of the variety of tasks in which the relationship had been observed. And because some cognitive operations seem to require more mental energy than others do, researchers attempted to link those processes to the concept of effort (see Jacoby et al., 1979; Nairne, Puse, & Widner, 1985). Similarly, the purpose of the research I describe here was to determine whether the generation effect in free recall is aligned with differences in cognitive effort.

The term *generation effect* refers to evidence that the cognitive act of generating, when compared with the act of reading the same information, produces an advantage for later remembering (see Jacoby, 1978; Slamecka & Graf, 1978). Our intuitions tell us that generating seems to be a more effortful task than reading, on the average—much as recall usually *seems* more effortful than recognition. If generation and effort are aligned, then conditions that produce an advantage of generating on an index of subsequent remembering should produce a corresponding impact from generating on an index of the amount of effort required by the initial orienting task. As described below, the effects of generating on indices of

effort and on subsequent recall were compared in four different contexts.

In the orientation phase of the experiment, anagrams and incomplete sentences were provided as contexts for decision making. The difficulty of solving the anagrams and completing the sentences was varied (easy vs. difficult) to constitute the four contextual conditions for each subject. Subjects in the *generating* condition viewed partial words (letters and blanks) and generated the complete words as solutions to the anagrams or the incomplete sentences by using rules on which they had been trained. Subjects in the *verifying* condition saw complete words and decided whether they were solutions to the corresponding anagrams or fit into the incomplete sentences. The verifying condition differed from the usual reading condition in the typical generation paradigm. The element of verification was included to provide a rationale for presenting the subjects with anagrams and incomplete sentences. Its inclusion, however, provides a good test for an effort account of the generation effect, because the two conditions (generation and verification) both require the subjects to relate words to their contexts.

A third type of processing task, *evaluating*, was invented in an attempt to vary effortful processing without having subjects generate the words. It was designed as a potentially effortful task, to which generation could be compared. After having been trained on the rules that were used to relate the words to their contexts, the subjects in this condition read the complete word, saw the context, and decided which rule had been used. Pilot subjects judged these evaluations to be more difficult than the verifications.

The inclusion of the evaluating condition in the design meant that more than one rule had to be used in each context. Accordingly, rules were chosen to correspond to the difficulty manipulation (see Tyler et al., 1979). Easy and difficult anagrams differed according to the number of letters in the word that were rearranged (one vs. four). Easy and difficult sentences differed according to the probability of producing the word from the incomplete sen-

I thank Richard Helmer, who helped program the tasks, and Ian Begg, Elizabeth Glisky, Reed Hunt, and Larry Jacoby, who commented on an earlier version of the manuscript. Please address requests for reprints to Paula Hertel, Department of Psychology, Trinity University, 715 Stadium Drive, San Antonio, TX 78284.

tence (1.00 vs. .04). Words were the first to come to mind to complete the easy sentences ("They hurriedly visited the leaning ____ of Pisa."). The words fit sensibly into the difficult sentences but were rarely produced in an independent completion task ("They hurriedly stole the ____ of silver coins."). (The missing word is *tower* in both cases.) "Rules" corresponding to these categories of difficulty were formulated and are reported below.

The basic requirement of any attempt to test an effort account of a phenomenon is the use of an index that defines effort independently from performance on the memory test. The most frequently used index of effort is the degree of cognitive capacity expended in performing a primary task, as measured by latency to respond in a secondary tone-detection task. Longer latencies reflect greater amounts of attention or effort allocated to the primary task. Such a measure was employed in the present research. While subjects either generated words or made decisions about the words they read in the primary task, tones were presented via headphones; the subjects were instructed to press a button whenever they detected a tone. Only if the pattern of latencies in responding to the tone was mimicked by the results of the free recall test would the concept of cognitive effort help us understand the memorial advantage of generating or performing other operations.

METHOD

Overview

Words were processed in the contexts of incomplete sentences and anagrams; each type of context varied in difficulty. The orienting task required the subjects to verify that a presented word fit the context, to generate the word from letters and blanks to fit the context, or to evaluate which rule had been used to choose the presented word for the context. Concomitantly with the orienting task, the subjects detected brief tones presented over headphones. The last task was an unannounced test of free recall.

Subjects and Design

Twenty volunteers from lower division psychology courses were randomly assigned to each of the three orienting tasks: verification, generation, or evaluation. The within-subjects factors were the type of context for the word (anagrams vs. sentences) and the level of difficulty in each context (easy vs. difficult). All subjects saw 10 words in each combination of context and difficulty.

The 20 subjects in the verifying condition decided if the word presented at the beginning of a trial either contained the letters of the following anagram or could be meaningfully inserted into the blank space in the sentence. In the generating condition, a partial word was presented at the beginning of the trial (i.e., d_c_o_), and the task was to generate the corresponding complete word that solved the following anagram or fit meaningfully into the sentence. These subjects received two rules for generating words from anagrams and two "rules" for generating words from sentences (one for each level of difficulty within each context). The 20 subjects in the evaluating condition saw the complete word at the beginning of each trial, as did those in the verifying condition, but they were required to decide which rule had been used to produce the word from the context.

Materials

Words and contexts. Nouns high in frequency, concreteness, and meaningfulness were selected from those used by Tyler et al. (1979) and were submitted to a pilot investigation in order to determine the probability of producing the words from the incomplete sentences. With

three productions per sentence required, the resulting 40 words were always produced to fill in the easy sentences; the probability of their production in the difficult sentences was .04. Anagrams for the same words were constructed by reversing the procedure described in the rules given below.

The 40 words were grouped into blocks of 10. Each block contained 4 five-letter words, 4 six-letter words, 1 seven-letter word, and 1 eight-letter word. Alternating blocks of words were assigned to each type of context—anagram and sentence—and the type of context in the first block was counterbalanced across subjects.

Rules. The following rules for unscrambling anagrams and filling in sentence blanks were typed on 3 × 5 cards, presented to subjects in the generating and evaluating conditions during the instructional period, and maintained in view. The anagram rules were: (1) move the first letter to the last position, (2) switch Letters 1 and 3 and switch the last two letters, and, in the evaluating condition only, (3) the letter string cannot be unscrambled to make a word. The sentence rules were: (1) the word is clearly the most sensible word to complete the sentence, (2) the word fits sensibly, although there are other words which would fit better, and, in the evaluating condition only, (3) the word does not complete the sentence sensibly.

Procedure

Orienting trials. The orienting task was implemented on a TRS-80 microcomputer, programmed to rotate all 40 experimental words through all conditions of context, difficulty, and tone delay. (The program began with the presentation of four practice trials that represented each context, level of difficulty, and tone delay.) Each trial began with the presentation of a complete or partial word at the top of the screen. Partial words contained letters in the first, third, and fifth positions, with the remaining positions represented by dashes. One second after the onset of the word (or partial word), the context appeared below it. The word (or partial word) and context remained displayed for 8 sec. On 80% of the trials, a weak but detectable tone occurred at delays of 1, 2, 3, or 4 sec after the onset of the context. The order of delays was randomized within blocks of 10 trials and fixed across subjects; the two levels of difficulty of the context were counterbalanced with tone delays within these blocks.

With the offset of the context, a question mark signaled the subject to say the word alone in the generating condition, or the word followed by the decision in the verifying or evaluating conditions. The subject pressed the "enter" key to terminate the trial. A blank screen lasting 1 sec separated trials.

Instructions. The general nature of both the primary and secondary tasks was described similarly to all subjects. They were asked to press the button (held in the nondominant hand) as quickly as possible in response to the tone, but without sacrificing attention to the primary task.

Instructions for the primary task varied according to the orienting conditions. The subjects in the verifying condition were told to verify that the word fit into the sentence sensibly or could be made from the anagram. They were asked to indicate each decision by repeating the word aloud and saying "yes" or "no," after the anagram or the sentence disappeared from the screen. These subjects were also informed that most trials would require a "yes" response, but that some subjects would experience a few negative instances; therefore, they should consider all decisions carefully.

The instructions for the generating subjects described the task to generate a complete word from a partial word by meeting the requirements of either one of the two rules for each context. The rules were described and examples provided. The experimenter administered four training trials for anagrams, using 3 × 5 cards. No overt decision about the rule was required, but the subjects were asked to delay the oral report of the word until the anagram or the sentence disappeared.

In the evaluating condition, instructions guided the subjects to expect the presentation of a complete word, followed by either context. They were trained on the rules for each context and instructed to decide which rule had been used to create the word. When the anagram or the sentence disappeared, they reported aloud the word and the choice of a rule.

Recall task. A short list of multiplication problems occupied the subjects during a 5-min retention interval. Subsequently the experimenter requested the free recall of at least 20 words. Guessing was encouraged if and when the subject expressed an inability to recall 20.

RESULTS

Recall

The mean proportions of the words recalled from those repeated or generated at the end of each trial are represented in Table 1. A mixed-design analysis of variance, with orienting task as the between-subjects factor and context and difficulty as the two within-subjects factors, revealed the following reliable effects, with the significance level set at .05.

First, the proportion of words recalled varied according to the type of orienting task [$F(2,57) = 8.49$, $MS_e = 0.039$]. A planned comparison revealed that generating produced better recall than verifying [.40 vs. .27; $F(1,57) = 15.21$]. Further analyses indicated that this difference occurred for words processed in each context independently [anagrams, $F(1,57) = 6.59$, $MS_e = 0.030$; sentences, $F(1,57) = 12.61$, $MS_e = 0.032$]. Verification and evaluation did not produce reliably different means in either context.

Difficult contexts produced overall better recall than did easy contexts [.36 vs. .28; $F(1,57) = 20.34$, $MS_e = 0.016$]. The interaction of difficulty with context was reliable [$F(1,57) = 5.98$, $MS_e = 0.023$]. Inspection of Table 1 reveals that the effect of difficulty was larger for words processed in the context of sentences; the effect was reliable only in that context [$F(1,57) = 26.32$, $MS_e = 0.017$].

Finally, any comparisons involving the generating condition in the difficult sentence context must be interpreted cautiously; a 29% error rate in generating the words from difficult incomplete sentences was obtained. (In contrast, the error rate in repeating or generating words, averaged across all other conditions, was 2.4%.) When recall was based on the proportion of all 10 words, the generating condition recalled an average of .36 from difficult sentences. The more conservative analyses revealed that the generation effect within the sentence context and the context \times difficulty interaction were not reliable.

Latencies

Table 2 reports the median response time in centiseconds within each context and level of difficulty, excluding trials in which the word was not repeated or generated. (Analyses performed on medians for all eight tone trials in each condition revealed patterns of significance identical to those reported below.) Failures to respond to the tone were rare (<1%).

Table 1
Mean Proportion of Words Recalled

Orienting Task	Context			
	Anagram Solutions		Sentence Completions	
	Easy	Difficult	Easy	Difficult
Verification	.22	.27	.26	.34
Generation	.35	.34	.36	.53
Evaluation	.23	.27	.29	.40

Note— $n = 20$.

Table 2
Mean of the Median Latencies in Centiseconds

Orienting Task	Context			
	Anagram Solutions		Sentence Completions	
	Easy	Difficult	Easy	Difficult
Verification	59	64	58	64
Generation	62	76	59	72
Evaluation	67	78	70	82

Note— $n = 20$.

The mixed-design analysis of variance revealed two reliable main effects and no reliable interactions. First, latencies in the secondary task varied according to the type of orienting task [$F(2,57) = 3.54$, $MS_e = 985.830$]. Planned comparisons between the verifying condition and each of the other conditions revealed that latencies in the evaluating condition were reliably longer than in the verifying condition [74 vs. 61; $F(1,57) = 7.07$, $MS_e = 985.830$], but that latencies in the generating and verifying conditions did not reliably differ (67 vs. 61). Further analyses showed that the difference between verifying and evaluating conditions held for each context separately [anagrams, $F(1,57) = 6.04$, $MS_e = 435.664$; sentences, $F(1,57) = 4.33$, $MS_e = 1028.282$].

A reliable main effect for the difficulty of the context was also obtained [$F(1,57) = 25.62$, $MS_e = 257.370$]. The more difficult contexts produced longer latencies in the secondary task (73 vs. 62 csec), replicating Tyler et al.'s (1979) findings with similar materials. Although the differences were small, they were quite stable, and they occurred within most orienting tasks and contexts, independently. Five of the six comparisons between easy and difficult conditions produced reliable effects. The effect was not found following verification in the anagram contexts.

DISCUSSION

According to the results of this experiment, the advantage of generating for subsequent recall is not related to effortful processing. The generating condition produced better recall overall than did the verifying condition, but not reliably longer latencies. When generating and verifying are compared within each of the four contexts (easy and difficult anagrams and easy and difficult sentence completions), it is clear that most contexts that produced the smaller differences in latencies (easy tasks) also produced the larger effects on free recall; and when the conservative measure of recall was employed, the generation effect was negligible for difficult sentences (.34 vs. .36).

Apart from their relevance to generation, the results of this experiment are interesting because they provide a clear instance of the failure of cognitive effort to uniformly predict memory performance. In this regard, the results from the evaluating condition are germane: The effect of evaluating rules on recall, when compared to verification, was not reliable in either context; the effect on latencies, however, was reliable in both contexts. Perhaps, however, a prerequisite for observing a correspondence of recall and effort is that effort must be expended on the material to be recalled, and not merely on information that is used to derive that material. This prerequisite was not clearly met in the instances that failed to show an effect of difficulty in this report.

As an illustration of the reasoning behind the suggested prerequisite, consider that Tyler et al. (1979) found an effect of difficulty in solving anagrams only when the word and the anagram were presented successively, and not when they were presented simultaneously, as in the present

experiment. Successive presentation requires that the word be held in working memory and compared with the anagram; simultaneous presentation allows decisions to be made without a complete activation of the word as a unit in memory, by checking for the presence of letters in both displays. In the present experiment, such activation was ultimately ensured by requiring subjects to repeat the word at the end of the trial. However, this procedure did not guarantee that the word was activated while the manipulation of difficulty was in effect (and verified by probes). At that point, subjects might have used a letter-checking procedure, then read the word for the purpose of reporting it at the end of the trial. The point is that the effortful processing of letter configurations might benefit the memory of those configurations, but not of the word.

Similarly, the effect of difficulty on recall was not clearly obtained within the generating condition, in which the word was produced as a final outcome of the trial (see Zacks, Hasher, Sanft, & Rose, 1983, for a similar finding). Again, better memory for the difficult contexts that produced the word might obtain, but better memory for the word is not guaranteed.

This line of reasoning can be extended to the observed dissociation between effort and generation. If effort must be allocated to the word that will be remembered in order to predict such memory, then its dissociation with the generation effect is not surprising. In the condition of verifying that words fit into sentences, the word is kept in mind while the decision is in process and the index of effort recorded. But in the generating condition, further processing is not required beyond the point of generating the word. In this analysis, the orienting conditions are naturally confounded by the focus of attention—on the word and its context, as opposed to the process of deriving the word from the context.

The present formulation can also be used to understand the lack of correspondence between effort and recall when the subjects' task was to evaluate rules. These evaluations appeared to require cognitive effort beyond that expended in verification, but perhaps the effort was not dedicated to processes that benefit later recall. This suggestion follows the reasoning put forth above regarding the simultaneous presentation of words and anagrams. Although the effect of evaluating words in the context of sentences was not reliable, there was a trend in that direction, especially with respect to the difficult sentences.

In conclusion, the present findings make it clear that the cognitive effort allocated to processing does not consistently predict memory performance. Several investigators of effort and memory (see Jacoby et al., 1979; Mitchell & Hunt, 1989) have suggested that any benefit for later recall must ultimately rely on the memorial consequences of the specific operations performed, rather than on the fact that effort has been expended. The argument conforms to our intuitions that it's not how hard you think, but *how* you think that has consequences for the future. Before the concept of processing effort is abandoned in the search for corre-

lates of remembering, however, it should be evaluated in terms of memory for the focus of its force.

REFERENCES

- BRITTON, B. K., WESTBROOK, R. D., & HOLDREDGE, T. S. (1978). Reading and cognitive capacity usage: Effects of text difficulty. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 582-591.
- EYSENCK, M. W., & EYSENCK, M. C. (1979). Processing depth, elaboration of encoding, memory stores, and expended processing capacity. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 472-484.
- GRIFFITH, D. (1976). The attentional demands of mnemonic control processes. *Memory & Cognition*, *4*, 103-108.
- JACOBY, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, *17*, 649-667.
- JACOBY, L. L., CRAIK, F. I. M., & BEGG, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, *18*, 585-600.
- JOHNSTON, W. A., & UHL, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 153-160.
- KELLOGG, R. T. (1984). Cognitive effort and memory revisited. *Psychological Reports*, *54*, 850.
- KRINSKY, R., & NELSON, T. O. (1981). Task difficulty and pupillary dilation during incidental learning. *Journal of Experimental Psychology: Human Learning & Memory*, *7*, 293-298.
- MITCHELL, D. B., & HUNT, R. R. (1989). How much "effort" should be devoted to memory? *Memory & Cognition*, *17*, 337-348.
- NAIRNE, J. S., PUSEN, C., & WIDNER, R. L., JR. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition*, *13*, 183-191.
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 592-604.
- TYLER, S. W., HERTEL, P. T., MCCALLUM, M. C., & ELLIS, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 607-617.
- WALKER, N., JONES, J. P., & MAR, H. H. (1983). Encoding processes and the recall of text. *Memory & Cognition*, *11*, 275-282.
- ZACKS, R. T., HASHER, L., SANFT, H., & ROSE, K. C. (1983). Encoding effort and recall: A cautionary note. *Journal of Experimental Psychology: Human Learning & Memory*, *9*, 747-756.

(Manuscript received April 25, 1989.)