Journal of Experimental Psychology:
Learning, Memory, and Cognition
1983, Vol. 9, No. 4, 747–756

Encoding Effort and Recall: A Cautionary Note

Rose T. Zacks
Michigan State University

Lynn Hasher, Henrianne Sanft, and Karen C. Rose
Temple University

In the five experiments reported here we attempted to demonstrate an effect on item memorability of the amount of effort expended during the encoding process. The encoding task in two experiments was anagram solving; here, solution difficulty was varied. In the third experiment, subjects were required to judge whether a word fit meaningfully into a sentence frame, and the ease of making this decision was manipulated. The final two experiments involved picture naming under time pressure; pictures were displayed either with no labels (easy condition) or, as in the picture–word version of the Stroop task, with superimposed interfering labels (hard condition). In none of the experiments did our manipulations of difficulty/effortfulness of encoding influence item retention. These findings raise questions about the robustness of the effort phenomenon.

The relation between amount of cognitive effort expended during encoding and the subsequent memorability of inputs has received considerable attention of late. A frequently reported finding is that the more effort required by an orienting or cover task, the greater is the incidental retention of the items encountered during the cover task (e.g., Jacoby, Craik, & Begg, 1979; Kolers, 1973, 1975; Krinsky & Nelson, 1981; Tyler, Hertel, McCallum, & Ellis, 1979). Given the theoretical and practical importance of this "cognitive effort" phenomenon, its reliability and boundary conditions are worthy of attention. Here, we report five experiments that used three different means of varying effort. In each, a surprise free-recall test examined the relation between effort and memorability.[1] Our findings suggest limits on the generality of the effort effect and indicate, as well, a need to determine the boundary conditions for the phenomenon.

This research was supported by National Institute of Mental Health Grant MH33140. The authors thank Aileen Mayzell for her assistance with the research. A review by T. O. Nelson of an earlier version of this article was also very helpful. Experiment 3A was carried out while R. T. Z. was on leave at Northwestern University.

Requests for reprints should be sent to Rose T. Zacks, Psychology Department, Michigan State University, East Lansing, Michigan 48824.

Experiments 1A and 1B

A major demonstration of the effort effect is found in a series of experiments by Tyler et al. (1979). One of the tasks used by these authors required subjects to judge whether a target word (or which of two target words) was the solution of an anagram. Difficulty was varied by using as easy anagrams ones in which there were few letter-order changes (from the target word) and as hard anagrams ones in which there were many letter-order changes. On a subsequent surprise recall test, hard anagram words were better recalled than easy ones. Tyler et al.'s judgment procedure was not exactly followed in the present research; instead, subjects in Experiments 1A and 1B actually solved anagrams that varied in their normative solution difficulty. A surprise recall

[1] Our original purpose in embarking on this series of studies was to test an implication of our view on encoding operations (presented by Hasher & Zacks, 1979). Specifically, we had hoped to show that in contrast to previous experiments that had used memory tests sensitive to amount of effortful processing (e.g., free recall), amount of effort at encoding would not influence retention of frequency of occurrence information, an event attribute encoded by an automatic process. The prediction about frequency information was supported, but the prediction about recall was not, mitigating the importance of the frequency findings. Therefore, only the recall data are reported here.

test of the solution words was then administered. We anticipated that words whose anagram formats were more difficult to decode would be better represented in recall than words whose anagram formats were less difficult.

Another variable, present only in Experiment 1B, was presence versus absence of a 5-sec, unfilled interval between individual anagram problems. This manipulation was an attempt to follow up on a suggestion by Jacoby (1978) that difficult processing engenders more interest and greater arousal than easy processing. If so, then subjects given extra time to think about the solutions to the anagrams would be more likely to reflect on the more "interesting" hard anagrams than on the easy ones. Thus a greater difference between easy and hard anagrams would be expected with the 5-sec interitem interval than with no interitem interval.

Method

Subjects. Subjects were introductory psychology students who participated to fulfill a course requirement. There were 32 subjects in Experiment 1A and 16 in each of the six groups of Experiment 1B. In these and all subsequent experiments, subjects were tested individually.

Design. In Experiment 1A, two levels of anagram difficulty (easy and hard) were crossed factorially with three levels of frequency of occurrence (1, 2, and 3) in a 2 × 3 within-subjects design. Experiment 1B included only once-presented items.[2] For two groups, anagram solution difficulty was varied within subjects; one of these received 5-sec intervals between anagrams, whereas the other did not. For the other four groups, both variables (presence vs. absence of rehearsal intervals and hard vs. easy anagrams) were varied between subjects.

Materials. Each subject was exposed to anagrams based on 18 different words selected from a total pool of 24 words. The words were ones whose letters could not be rescrambled to make any other English word (i.e., single-solution anagrams; Olson & Schwartz, 1967).

The normative data of Mayzner and Tresselt (1958) were used to generate the easy and hard anagrams. They had found that when the letter orders of anagram and word were maximally similar (easy letter orders), the solution was quickly obtained; by contrast, when the letter orders of anagram and word were maximally dissimilar (hard letter orders), the solution process was more prolonged. Three unique easy letter orders and three unique hard letter orders were chosen. In Experiment 1A, words occurring two and three times were shown in a different letter order at each presentation. For words occurring one and two times, the particular letter orders used were selected randomly from the three available at each difficulty level. In Experiment 1B, each unique letter order was used with an equal number of subjects.

Throughout this series of experiments, multiple lists were constructed from the basic set of materials. These functioned to counterbalance the items across the difficulty levels and (where relevant) across the frequency levels. In Experiment 1A, there were eight unique lists consisting of 36 anagrams with each list given to four subjects. Lists of 18 anagrams were used in Experiment 1B. In the within-subject case, four unique lists were used with each given to four subjects per group. In the between-subject case, eight easy and eight hard lists were generated with each given to two subjects.

The general procedure for generating presentation orders was as follows: Repetitions of items receiving multiple presentations (e.g., in Experiment 1A) were always spaced, and items representing different conditions (e.g., easy vs. hard encoding condition) were evenly distributed across input positions. Otherwise, presentation order was random.

Procedure. Each anagram was typed on a separate 3 × 5-in. (7.6 × 12.7 cm) card. Except in the rehearsal-interval conditions of Experiment 1B where a 5-sec wait was instituted between cards, subjects turned to the next anagram as soon as the current one was solved or at the end of 2 min if the anagram remained unsolved. After 90 sec of unsuccessful work on an anagram, subjects were given the first letter of the solution word to use as a hint. Subjects reported their solutions aloud, and solution times were recorded. Following the final anagram, an unexpected free-recall test of the solution words was administered. A maximum of 3 min was allowed for written recall.

Results and Discussion

Experiment 1A. An alpha level of .05 was used throughout. Solution probabilities and solution times were used as indices of the relative difficulty of the two types of anagram solution rules.[3] Words receiving multiple presentations were always solved after the first one, but solution probability did differ on the first presentation with easy anagrams more frequently solved (.99) than hard ones (.94), $t(31) = 5.00$. Solution times also indicated an

[2] This is the typical procedure in experiments on effort effects. Our other experiments included frequency of occurrence as a variable because of their initial purpose (see Footnote 1). Experiment 1b was run last and includes manipulations that we thought would maximize the opportunity for finding effort effects on recall.

[3] It should be acknowledged that these measures have problems. Solution probabilities showed little variance, and response time measures have been critized as indicants of cognitive efforts (e.g., Krinsky & Nelson, 1981; Tyler et al., 1979). It is perhaps unfortunate that another measure of effort such as performance on an appropriate secondary probe task (Tyler et al., 1979) or pupil diameter (Krinsky & Nelson, 1981) was not included in the present studies. At the time, however, we did not think this necessary; the effort phenomenon seemed well established, and at least for Experiments 1A, 1B, and 2, we used variants of Tyler et al.'s (1979) procedures. These authors had validated the effort manipulations by secondary-task performance.

Table 1
Experiment 1A: Mean Solution Times (in sec) and Recall Probabilities at Each Presentation Frequency and Difficulty Level

Difficulty	No. of presentations 1	2	3	M
Solution times				
Easy	13.93	6.49	2.32	7.58
Hard	37.55	13.66	7.52	19.58
M	25.74	10.08	4.92	—
Recall				
Easy	.28	.45	.80	.51
Hard	.32	.48	.74	.51
M	.30	.46	.77	—

effective manipulation of difficulty (see Table 1). Hard anagram rules resulted in longer solution times than easy anagram rules, $F(1, 31) = 20.45$, $MS_e = 337.71$. Solution times declined with repetition, $F(2, 62) = 82.89$, $MS_e = 67.44$, and they declined to a greater extent for hard than for easy items, $F(2, 62) = 24.25$, $MS_e = 67.44$.

Recall scores are also shown in Table 1. As expected, the greater the number of presentations, the greater the recall, $F(2, 62) = 73.16$, $MS_e = .45$. However, the difficulty dimension did not influence recall either as a main effect or in interaction with the number of repetitions an item had (*F*s < 1). Even the largest difference between easy and hard items, that for items presented only once, was not significant, $t(31) = 1.25$, $p > .05$. These conclusions are unaltered by considering recall for only those anagrams that were actually solved; for example, for once-presented items, the proportions of easy and hard words recalled were, respectively, .26 and .29, $t(31) = .43$.

Experiment 1B. The results of the two groups for which difficulty was a within-subjects variable are considered first. For these groups, both measures of performance on the anagram task, solution probability and solution time, are consistent in showing no effect of pacing interval on solution performance but a significant effect of anagram difficulty. Hard anagrams were solved less frequently than easy ones (solution probabilities of .89 and .98, respectively), $t(31) = 2.92$; and, as can be seen in Table 2, they took longer to solve than did easy ones $F(1, 30) = 17.57$, $MS_e = 261.13$. Despite this apparently effective manipulation of solution difficulty, recall of easy versus hard anagrams did not differ for these two groups ($F < 1$). In fact, the analysis of the recall data revealed no significant effects (*F*s < 1 for the main effect of presence vs. absence of rehearsal interval and for the Interval × Difficulty interaction).

The results for the four groups for which solution difficulty was varied between subjects are similar. Hard anagrams were solved at a lower rate (.87) than were easy anagrams (.98), $t(62) = 145.38$; and hard anagrams took longer to decode than easy ones, $F(1, 60) = 43.24$, $MS_e = 292.79$ (see Table 2).

Once again rule difficulty failed to influence recall, $F(1, 60) = 1.20$, $MS_e = 5.73$. Also, the

Table 2
Experiment 1B: Mean Solution Times (in sec) and Recall Probabilities at Each Difficulty Level, Pacing, and Design Condition

Difficulty	Difficulty within-subject: No delay	5 sec delay	M	Difficulty between-subject: No delay	5 sec delay	M
Solution times						
Easy	17.10	18.54	17.82	14.24	17.46	15.85
Hard	37.36	32.14	34.75	46.61	41.36	43.98
M	27.23	25.34	—	30.42	29.41	—
Recall						
Easy	.36	.38	.37	.43	.40	.42
Hard	.38	.33	.36	.34	.41	.38
M	.37	.35	—	.39	.41	—

group with additional time to reflect showed no advantage in recall ($F < 1$), and the interaction between solution difficulty and pacing condition was not significant, $F(1, 60) = 2.29$, $MS_e = 5.73$. Analyses performed on the recall of only those anagrams actually solved produced identical outcomes.

In trying to understand our failure to find any effort effects in several comparisons between easy and hard anagrams, we reconsidered Tyler et al.'s (1979) anagram data. These authors found an effort effect only in one of three experiments, the one in which successive presentation of target word and anagram was instituted in order to prevent subjects from using a simple matching strategy such as was possible with the simultaneous presentation procedure of the other two experiments. Because the current experiments required the subjects to generate solutions for the anagrams, no noneffortful strategy was feasible, and a difficulty effect in recall seemed a reasonable expectation.

Do our results constitute a failure to replicate Tyler et al.'s research with anagrams, particularly the one experiment in which an effort effect was obtained? We thought we had incorporated the essential features of that experiment when we designed our studies. However, after the fact, a number of possibly crucial differences between our procedures and theirs can be suggested. One such difference, mentioned by Ellis (Note 1) relates to the fact that in Tyler et al.'s research, the target items were presented at the beginning of each encoding trial and thus available (physically or in memory) while the encoding task was being performed. In our experiments, by contrast, the target words were generated from the anagrams and thus available only after the encoding task had been completed. The importance of this and other differences in procedure can only be established through future research. In any case, this particular difference does not apply to Experiment 2, which followed Tyler et al.'s method with regard to the sequencing of events during encoding.

Experiment 2

Because we could find no relation between encoding effort and recall using an anagram task, we decided to switch tasks. Now we took a task in which solution difficulty had already been demonstrated to influence recall in three experiments (Tyler et al., 1979). In this task, subjects are shown a target word and a sentence frame that is missing a single word. The subject decides whether the target word meaningfully completes the sentence. Decision difficulty is varied by having as easy sentences ones in which the missing word is highly likely to be one particular word and as hard sentences ones in which the missing word is not so readily guessed.

Method

Design and subjects. The design was a 2 (easy vs. hard decisions) × 2 (yes vs. no answers) × 3 (frequencies 1, 2, and 3) within-subject factorial. The free-recall test of the target words was a surprise. Twenty-four undergraduate students were each paid $2.00 for their participation.

Materials. Thirty-six common words were chosen from the Thorndike–Lorge (1944) norms. For each of them, six sentences were written to serve as meaningful frames (i.e., the correct response to the question of whether the word fits in the frame is yes); three were in an easy format and three in a hard format. Most often, hard format sentences were created by removing a concrete referent having some preexisting associative strength to the target word. For example, for the target word *doctor,* one of the easy/yes frames was "The emergency room was chaotic because there was no ___ available." In the hard/yes version of this frame the word *emergency* was deleted.

To validate the materials, we had a separate group of 15 subjects rate them as to the ease with which a target word completed a sentence frame. Recently, Ellis (Note 2) has provided data suggesting that such ratings are valid measures of amount of effort required for sentence-completion tasks: He has found that ratings correlate highly with effort estimates based on secondary task reaction times. In the rating forms for the current experiment, each target word was shown with the parallel easy and hard versions of the same basic sentence frame (e.g., the word *doctor* appeared with the two frames just described). The raters were asked to choose the frame in each pair that more clearly went with the target word. The frames that we had designated as easy ones were selected 89.6% of the time, thus indicating (in conformity with Ellis's data) that the materials would provide an effective manipulation of decision difficulty in the case of the *yes* decisions.

Our materials also included anomalous frames (i.e., ones that are not meaningfully completed by any target words and for which the correct answer on the judgment task is therefore *no*). However, neither the ratings obtained in the rating procedure just described nor the decision latencies obtained in the main experiment indicated that the anomalous sentences allowed an adequate manipulation of decision difficulty. Therefore, no further mention is made of the data from the *no* conditions.

The initial pool of 36 words was divided into four sets of nine words each. These were rotated among the four basic experimental conditions (easy vs. hard × yes vs. no). Words were also rotated across frequency levels (1, 2, and

3 presentations) and the result was 12 different lists each of which was given to two subjects. With 2 buffer items at the beginning and 2 at the end of the list, each list's total presentation series consisted of 76 items. When target words appeared more than once, each repetition was paired with a different sentence frame.

Procedure. The apparatus consisted of an AIM 65 microcomputer interfaced with a slide projector and a response box containing two operative response buttons, the left one for *no* answers and the right one for *yes* answers. On each presentation trial, a blank slide was presented for 1,500 msec, followed by a target word presented for 1,500 msec, followed by a a sentence frame. The subject terminated the sequence by pressing the appropriate response button. The computer controlled the slide projector and recorded the subject's responses as well as the response latencies.

Instructions for the judgment task asked subjects to decide whether each target word could complete the sentence frame with which it was paired. Both speed and accuracy of response were emphasized. Subjects were also informed that some of the target words would occur more than once. An unpaced, unexpected, written free-recall test of the target words immediately followed the judgment task.

Results and Discussion

Mean decision times for *yes* responses are shown in Table 3. These data are based on all trials including the 5% of trials on which the subjects disagreed with our determination that the target word fit the sentence frame. Probably because the sentence frame was different on each repetition of a word, decision time did not decrease over presentations for repeated targets, $F(2, 46) = 1.22$, $MS_e = 110{,}763$. Also, unexpectedly, hard and easy decisions did not differ in response time, $F(1, 23) = 2.11$, $MS_e = 153{,}890$. This latter finding is probably due to the fact that the time measure included sentence reading time, and our easy frames tended to be longer than our hard ones (means of 10.42 and 8.80 words for easy and hard frames, respectively). The data from the rating group indicate that we did have an effective manipulation of difficulty/effort in the *yes* condition. Also, our materials closely modeled those of Tyler et al. (1979) who validated the difficulty manipulation by means of both reaction times on a secondary task and difficulty ratings.

Table 3 also presents the recall data. These provide no evidence that sentences requiring difficult decisions result in superior recall for their targets ($F < 1$). This is true even if one only considers *yes* sentences and then only for Frequencies 2 and 3 where there is no possibility of a floor effect. The only variable that influenced recall was number of presentations, $F(2, 46) = 33.31$, $MS_e = .53$. Thus, once again we failed to find an effect of encoding effort.

Table 3
Experiment 2: Mean Decision Times (in msec) and Recall Probabilities (in %) at Each Difficulty and Presentation Level for the Yes Conditions

	No. of presentations			
Difficulty	1	2	3	*M*
Decision times				
Easy	3,336	3,166	3,242	3,248
Hard	3,140	3,116	3,212	3,156
M	3,238	3,141	3,227	3,202
Recall				
Easy	7	29	50	29
Hard	10	25	47	27
M	8	27	49	—

Experiments 3A and 3B

Our final two attempts at demonstrating a cognitive effort effect on recall used a picture–word version of the Stroop interference task. Subjects had to name pictures of familiar objects that in some cases had interfering verbal labels superimposed on them (e.g., a picture of a cigar with the word *pipe*) and in other cases did not (e.g., the picture of the cigar alone). As shown by response time measures, interfering labels make picture naming more difficult (e.g., Lupker, 1979). This presumably occurs because the interfering word is read automatically and the suppression of its utterance requires considerable cognitive capacity, which slows down naming the picture (e.g., Dyer, 1973). Thus, according to a simple version of the effort hypothesis the difficult labeled pictures should be better recalled than the easier unlabeled pictures.

Label condition was varied between subjects in Experiment 3A and within subjects in Experiment 3B. Also, in the latter experiment, some of the subjects were forewarned of the upcoming recall test.

Method

Design and subjects. Experiment 3A had a 2 (labeled vs. unlabeled pictures) × 3 (frequency levels 1, 2, and 4) factorial design with the label factor manipulated between subjects and the frequency factor within subjects. Twenty volunteer undergraduate subjects participated in each group.

The design of Experiment 3B was a 3 (label condition) × 3 (frequencies 1, 2, and 4) × 3 (instructions) factorial with only the last factor tested between subjects. The three label conditions were no labels, labels unrelated to the picture they were superimposed on, and related labels. (In some cases at least, related labels have been found to make picture naming more difficult than unrelated labels—see Lupker, 1979; if so, all three label conditions should vary in recall.) A within-subject design was used because in it each subject would experience the various levels of difficulty. The three instructrional conditions included an incidental group informed only about the Stroop task, a standard intentional group told about the recall task and asked to treat both the recall and Stroop tasks as equally important, and an intentional group told that the recall task was more important than the Stroop task. Intentional memory instructions were included to enhance any weak effort effects that might be present; if differences in effort expended at encoding result in differential item availability, and if subjects forewarned about a subsequent test expend more of their rehearsal effort on the more available items, these items should have an advantage in recall. There were 16 undergraduate subjects in each group.

Materials. For Experiment 3A, the norms of Snodgrass and Vanderwart (1980) were used to select 36 line drawings that met two criteria: (a) Each had received at least a 90% agreement level for the labels assigned, and (b) they represented a broad range of semantic categories. Interfering labels for each picture were generated in a pilot study using 10 undergraduates. Each was shown the pictures one at a time and was asked to generate the first word they thought of (other than the name of the picture) from the same semantic category. For each picture the modal response was chosen to serve as the interfering label in the experimental task. In each label condition, there were four unique lists, each containing nine pictures at each frequency level. The presentation series contained 72 pictures, the 1st and last 3 of which were buffer pictures. For pictures presented more than once, the labels were always the same.

The 36 critical pictures of Experiment 3B were line drawings taken from the Peabody Picture Vocabulary Test. The related labels for these pictures were determined by searching through association norms and by asking eight people to give free associations (nouns) to a list of nouns that named the experimental pictures. Unrelated labels were chosen so as not to be related to any of the critical pictures. Four different lists each contained three instances of each of the nine (Label × Frequency) within-subject conditions. The presentation order for each list contained 80 pictures that included five buffers at the beginning and end of the list and seven fillers spread throughout.

Procedure. The equipment included an AIM 65 microcomputer interfaced with a slide projector and a voice-activated relay. The computer timed subjects from the onset of each slide to their naming of the picture. There was a 1,000-msec delay between the naming of a picture and the presentation of the next one. Subjects were instructed to name each picture as quickly and accurately as they could. They were warned not to correct out loud any errors they might make but instead to wait for the next slide. Before the experiment proper began, subjects were given practice at responding into the microphone by having them read a series of 20 slides of first names.

Only subjects in the two intentional conditions of Experiment 3B were forewarned of the recall test. The standard intentional subjects were simply asked to "work at remembering the pictures while doing the naming task." The subjects for whom recall was emphasized were told that the purpose of the experiment was to investigate memory for pictures and that although the naming task was not trivial, their basic objective was to remember the pictures. All subjects were given a written, unpaced, free-recall test of the pictures.

Results and Discussion

Experiment 3A. As found in previous research, the group with interfering labels took longer than the group with no superimposed labels to name the pictures, $F(1, 38) = 16.72$, $MS_e = 45{,}005.92$ (see Table 4). For both label conditions, picture-naming latency decreased as the number of presentations increased, $F(3, 114) = 29.28$, $MS_e = 4{,}801.71$.

The recall data (also shown in Table 4) revealed no effort effect; easy items (without a label) were as well recalled as hard items ($F < 1$).[4] Recall did improve with frequency, $F(2, 76) = 28.93$, $MS_e = .71$. The almost significant interaction between frequency and label condition, $F(2, 76) = 2.75$, $MS_e = .71$, $p = .07$, subsumes the one finding that favors the effort hypothesis: For pictures presented four times, there was better recall of pictures shown with interfering labels. However, the equivalent difference in the opposite direction for once-presented items tempers the interpretation of this finding. Both differences were close to significance on one-tail tests, $t(38) = 1.59$ and 1.68 for items presented once and four times, respectively ($.05 < p < .06$).

Experiment 3B. Instructional condition was not involved in any significant main effects or interactions for either reaction time or recall. Therefore, the reported data are collapsed across instructional condition.

As in Experiment 3A, the label variable exerted a significant influence on reaction time, $F(2, 90) = 36.66$, $MS_e = 9{,}046.3$. As can be seen from Table 5 the difference resides largely

[4] It is possible that the lack of a difficulty effect in the Experiment 3A condition with interfering labels involves a trade-off between the positive impact of increased effort and the negative impact of the labels increasing memory load. This problem does not pertain to Experiment 3B in which label condition was varied within subjects.

Table 4
Experiment 3A: Mean Picture-Naming Times (in msec) and Recall Probabilities (in %) for Label and No-Label Conditions at Each Presentation Frequency Level

	No. of presentations				
Label condition	1	2	3	4	*M*
Naming Times					
No label	1,606	1,541	1,484	1,489	1,530
Label	1,758	1,655	1,619	1,635	1,667
M	1,682	1,598	1,552	1,562	—
Recall					
No label	41	50	—	60	51
Label	32	54	—	68	51
M	37	52	—	64	—

between having no label and having any label superimposed. The difference between a related and unrelated label was minimal, 1,416 and 1,417 msec, respectively (cf. Lupker, 1979). Also as in Experiment 3A, reaction times became shorter with repeated exposure to individual pictures, $F(3, 135) = 159.32$, $MS_e = 8{,}740.0$.

Mean recall levels are shown in Table 5. Recall increased with the number of repetitions an item had, $F(2, 90) = 20.85$, $MS_e = .52$. However, effort expended at encoding made little apparent difference in performance ($F < 1$). The only significant effect involving presumed effort differences was an interaction between the label conditions and frequency level, $F(4, 180) = 2.88$, $MS_e = .50$. Inspection of the means involved in this interaction (see Table 5) suggests that for once-presented items, there may have been a difficulty effect. However, the small difference in favor of the related labels condition was not significant on a Newman-Keuls test and only approached significance ($p < .10$) when an orthogonal contrast compared related labels with the average of the other two conditions. Even this weak support for the effort effect does not apply to pictures presented twice. Here results are in the opposite direction from those predicted, and a Newman-Keuls test (at .05) shows that pic-

Table 5
Experiment 3B: Mean Picture-Naming Times (in msec) and Recall Probabilities (in %) at Each Label Condition and Presentation Frequency Level

	No. of presentations				
Label condition	1	2	3	4	*M*
Naming times					
None	1,466	1,340	1,295	1,276	1,344
Unrelated	1,578	1,412	1,354	1,325	1,417
Related	1,556	1,401	1,363	1,343	1,416
M	1,533	1,384	1,337	1,315	—
Recall					
None	26	47	—	53	42
Unrelated	25	39	—	58	41
Related	33	33	—	52	39
M	28	40	—	54	—

Note. Data are collapsed over instructional condition.

tures seen with no labels were better recalled than those seen with related labels.

Again, there is no evidence for an effort effect: Recall for items that were difficult to process was no higher than for items easy to process. That the instructional manipulation also did not affect recall may perhaps be explained by the rapid pacing of the naming task (only a 1-sec delay followed the naming of each picture) and by the dual-task demands of naming the pictures as rapidly as possible while preparing for recall.

General Discussion

Our initial reading of the literature had suggested that the effort expended during encoding had a reliable effect on the ultimate retrievability of relevant memory episodes (e.g., Jacoby, 1978; Jacoby et al., 1979; Kolers, 1973, 1975). And so we began a series of studies designed to capitalize on this seemingly robust finding (see Footnote 1). In five attempts, we failed to find evidence of a reliable relation between encoding effort and recall. These findings suggest that (a) the effort effect should not be thought of as robust and (b) its boundary conditions need to be determined empirically.

It seems to us that one major methodological issue might prevent a reader from accepting our findings: We did not include any direct measure of the effort expended during encoding. For example, we did not repeat Tyler et al.'s use of a secondary task. Such tasks potentially provide "manipulation checks" for validating whether effort on the main task was actually varied. Our findings are thus open to the argument that we failed to find an effort effect because we failed actually to vary effort. However, for each of the experiments, a claim of effective manipulation of effort demands is supported by previous research. In Experiments 1A and 1B, we used a task (anagram solving) that had previously been used (Tyler et al., 1979) to demonstrate at least weak effort effects on recall, and we varied difficulty in accordance with normative data (Mayzner & Tresselt, 1958). The difficulty manipulation in Experiment 2 was taken from Tyler et al. (1979), and our materials were patterned after theirs.[5] Finally, with respect to the Stroop phenomenon used in Experiments 3A and 3B, considerable data and numerous theoretical discussions indicate that it takes less effort for a person to name an unlabeled picture than to name one shown with an interfering label (e.g., Dyer, 1973). Also, although the error patterns of Experiments 1A and 1B, the response time results of Experiments 1A, 1B, 3A, and 3B, and the sentence-frame/target-word compatibility ratings of Experiment 2 may not be compelling in and of themselves, they do provide some confirmation of adequate manipulation of effort.

Another issue is whether performance on a subsidiary task would have allowed clearcut assessment of the effort expended on the primary task. As Fisk, Derrick, and Schneider (Note 3) have recently noted, dual-task methodology poses numerous pitfalls. In fact, they argue that many dual-task experiments (including the work of Tyler et al., 1979) have failed to measure expended capacity accurately because one or more of three incorrect processing assumptions are reflected in the experimental methodology: (a) that processing capacity is undifferentiated with respect to the "structures" involved (e.g., input and output modes), (b) that subjects' strategies (e.g., which of two tasks to emphasize) do not influence the results, and (c) that performance does not change with practice during the tasks. In a dual-task experiment that took into account all of these factors, Fisk et al. found that increased encoding effort did not benefit recognition performance; in one instance there was even a trend in the opposite direction.

Elsewhere in the literature there are a number of other studies in which encoding effort was ostensibly varied that also failed to find any differences in recall. Consider Jacoby's (1978) findings with an encoding task in which subjects were given a cue word (e.g., *foot*) and were asked to generate a related response word from some of its letters (e.g., s _ _ e). Difficulty was varied by omitting one versus two letters of the response words. This affected the probability of response generation but did not influence cued recall of either once-presented

[5] It is unlikely that the absence of a secondary task in our research was a critical factor in our failure to find effort effects similar to those of Tyler et al. These authors' Experiment 1 did not include the secondary task but gave results similar to their other experiments.

response words or of massed twice-presented words. Similarly, Craik and Tulving (1975) reported that the difficulty (defined in terms of response time) of a nonsemantic encoding task had little influence on retention (see also Walsh & Jenkins, 1973).

One might want to argue that failures to find effort effects demonstrate that amount of effort is not the only important variable. For example, it could be claimed that effort expended influences recall only if effort manipulation involves "task relevant" processing. Such claims may well be valid, but in the absence of any precise definition of variables such as task relevance, they lack utility. That is, as it stands now, there appears to be no a priori way of knowing which "effort" manipulations will influence retention and which will not. Clearly, the need for further theoretical and empirical work on this interesting phenomenon is acute.

Our rereading of the literature on encoding effects also suggests that at this time it is useful to maintain a distinction between the effort effect and one that is often treated as related, the "generation" effect (Slamecka & Graf, 1978). Encoding conditions that require people to generate to-be-remembered items result in better retention than conditions in which people are provided with the same items. The crossword-puzzle-encoding task used by Jacoby (1978) allowed a demonstration of this effect; response words that subjects generated from clues had higher recall than did words that were presented complete and simply read by the subjects. Effort and generation effects are often treated as instances of the same phenomenon (e.g., Cuddy & Jacoby, 1982; Jacoby et al., 1979; Tyler et al., 1979), with the generation condition presumed to be more difficult than the nongeneration condition.

We recommend that these two effects be treated separately because the generation effect appears to be more robust than the effort effect. The former has been obtained with a wide variety of encoding and retrieval tasks (see Slamecka & Graf, 1978, and McElroy & Slamecka, 1982, for reviews).[6] By contrast, our research, as well as that of others (Craik & Tulving, 1975; Jacoby, 1978; Walsh & Jenkins, 1973; Fisk et al., Note 3), suggests that the scope of the effort phenomenon may be limited. Differences in the reliability of finding the effects, coupled with the absence of any well-articulated conceptualizations of either effect, suggest the importance of maintaining a distinction between the generation and effort effects.

[6] Thus far, the only factor that has been identified as eliminating a generation effect is the use of a generation task that results in a nonmeaningful generated response (McElroy & Slamecka, 1982).

Reference Notes

1. Ellis, H. Personal communication, September 1982.
2. Ellis, H. C. *Cognitive effort and mood states in memory.* Paper presented at the meeting of the Psychonomics Society, November 1982.
3. Fisk, A. D., Derrick, W. L., & Schneider, W. *The use of dual task paradigms in memory research: A methodological assessment and an evaluation of effort as a measure of levels of processing* (Research Report HARL-ONR-8105). Chicago: University of Illinois, Human Attention Research Laboratory, 1982.

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.

Cuddy, L. J., & Jacoby, L. L. When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Memory,* 1982, *21,* 451–467.

Dyer, F. N. The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition,* 1973, *1,* 106–120.

Hasher, L., & Zacks, R. T. Automatic and effortful processes in memory. *Journal of Experimental Psychology: General,* 1979, *108,* 356–388.

Jacoby, L. L. On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior,* 1978, *17,* 649–667.

Jacoby, L. L., Craik, F. I. M., & Begg, I. Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning and Verbal Behavior,* 1979, *18,* 585–600.

Kolers, P. A. Remembering operations. *Memory & Cognition,* 1973, *1,* 347–355.

Kolers, P. A. Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory,* 1975, *1,* 689–701.

Krinsky, R., & Nelson, T. O. Task difficulty and pupillary dilation during incidental learning. *Journal of Experimental Psychology: Human Learning and Memory,* 1981, *7,* 293–298.

Lupker, S. J. The semantic nature of response competition in the picture–word interference task. *Memory & Cognition,* 1979, *7,* 485–495.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of letter order and word frequency. *Journal of Experimental Psychology,* 1958, *56,* 376–379.

McElroy, L. A., & Slamecka, N. J. Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning and Verbal Behavior,* 1982, *21,* 249–259.

Olson, R., & Schwartz, R. Single and multiple solution five-letter words. *Psychonomic Monograph Supplements,* 1967, *2,* 105–152.

Slamecka, N. J., & Graf, P. The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory,* 1978, *4,* 592–604.

Snodgrass, J. G., & Vanderwart, M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory,* 1980, *6,* 174–215.

Thorndike, E. L., & Lorge, I. *The teachers' word-book of 30,000 words.* New York: Columbia University, Teacher's College, Bureau of Publications, 1944.

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory,* 1979, *5,* 607–617.

Walsh, D. A., & Jenkins, J. J. Effects of orienting tasks on free recall in incidental learning: "Difficulty," "effort," and "process" explanations. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 481–488.

Received August 15, 1982
Revision received January 4, 1983 ■