1979

Cognitive Effort and Memory

S. W. Tyler

Paula T. Hertel
Trinity University, phertel@trinity.edu

M. C. McCallum

H. C. Ellis

Repository Citation

Tyler, S.W., Hertel, P.T., McCallum, M.C., & Ellis, H.C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory, 5*(6), 607-617. doi: 10.1037/0278-7393.5.6.607

Cognitive Effort and Memory

Sherman W. Tyler, Paula T. Hertel, Marvin C. McCallum, and
Henry C. Ellis
University of New Mexico

We propose that the concept of cognitive effort in memory is both useful and important. Cognitive effort is defined as the engaged proportion of limited-capacity central processing. It was hypothesized that this variable might have important memorial consequences and might also be a potential confounding factor in levels-of-processing paradigms. The first experiment tested this possibility using two types of incidental-learning tasks factorially combined with two degrees of effort. It was found that high effort led to better recall than low effort, but that level-of-processing effects were nonsignificant. A second experiment clearly demonstrated the feasibility of using performance on a secondary task as an independent criterion for measuring effort, and two further experiments ruled out alternative accounts of effort effects. A reliable levels-of-processing effect was obtained in the fourth experiment in which the incidental-learning tasks were blocked. Implications and possible future applications of the cognitive effort concept are discussed.

The concept of cognitive effort can be best appreciated by examining it within the context of levels-of-processing research. The levels-of-processing approach has been an especially popular one in recent years, both as an impetus to research and as an explanatory device for many general findings in memory (cf. Craik & Lockhart, 1972). The original notion was based on the premise that a series of analyzers, varying along a continuum from structural to semantic analysis, is employed in processing an item for storage in memory; and the greater the depth to which an item is processed, the greater the probability it will later be recalled. Repeated processing

of a word to a given depth was assumed not to improve memory beyond the level associated with that depth.

This general view has been under considerable attack lately. For example, Nelson (1977) has argued that there are two major problems with this viewpoint. First, it can be demonstrated empirically that repeated processing of an item to the same depth does in fact improve memory relative to nonrepeated processing. Second, the notion of *depth of processing* is theoretically vapid, since there is no independent measure by which to order different processing tasks along the depth continuum; the best that can be done is divide the available tasks into two or three nominal categories using rather loose criteria. Baddeley (1978) has made similar arguments, adding the point that constructs later incorporated into this approach, including notions such as compatibility and breadth of processing, suffer from the same lack of any independent measurement criteria. Additional doubts about depth have been recently noted by Nelson, Walling, and McEvoy (1979).

Given this state of affairs, we asked if there could be an identifiable factor in

memory processing, the effects of which are directly reflected in recall performance, and which can be described as a continuum accessible to independent measurement. We propose that a reasonable candidate is mental or cognitive effort, or more explicitly, the amount of the available processing capacity of the limited-capacity central processor utilized in performing an information-processing task. The procedures for measuring effort by evaluating the attention-demand characteristics of stimuli under various task conditions were initially described by Ellis and Kreezer (Note 1). Their approach used a methodology based on divided attention, in which the subject performed a primary motor-tracking task and a secondary task involving simple reaction time to auditory or visual stimuli. The effort concept, more formally introduced by Moray (1967), seems particularly suitable in a number of ways. Not only has it found considerable utility in work such as that of Johnston and Uhl (1976), Johnston, Wagstaff, and Griffith (1972), Posner and Snyder (1975), Keele (1973), and Kahneman (1973), but it also is susceptible to independent measurement through the use of a secondary task, a procedure well reviewed by Kerr (1973). Furthermore, it appears logical to assume that the extent of involvement of an individual in a cognitive task would have a substantial impact on later memorial performance.

The concept of cognitive effort was proposed by Ellis (Note 2) as a mechanism for accounting for perceptual grouping effect in recall. In several studies (e.g., Ellis, Parente, Grah, & Spiering, 1975; Ellis, Parente, & Walker, 1974), letter strings that were variably grouped on successive presentations were recalled substantially better than letter strings presented in constant fashion. This finding, known as the variability effect in recall, is thought to be the result of varied input leading to greater effort in processing the letter strings than constant input. In general, recall is substantially greater when subjects reorganize the information rather than process it as presented by the experimenter. Similar findings have been reported more recently by Slamecka and Graf (1978) and by Jacoby (1978). Neither, however, provides evidence in support of an effort hypothesis. More specifically, Ellis (Note 2) proposed that the amount of effort in processing a given class of information can vary independently of the level at which it is processed. Thus, it is possible to process superficial structural (perceptual) information as well as semantic information with varying degrees of effort. Accordingly, the degree of cognitive effort is viewed as an alternative hypothesis in accounting for several memory phenomena.

The possibility that effort, as a factor separable from level of processing, can produce recall differences has not been totally ignored. Walsh and Jenkins (1973) did report a series of experiments designed to dismiss this possibility. Arguing that no independent definition of effort existed, they attempted to manipulate effort by requiring subjects to perform either one or two different incidental tasks on each stimulus item, the latter presumably requiring greater effort. Finding, in the two-task case, that the amount recalled was greater for the task requiring a deeper level of processing and otherwise did not differ from the one-task case, they rejected effort as an influential factor. The major problem with this study is the confusion of total time with momentary effort. Performing two tasks does not necessarily require a larger amount of central-processing capacity at a given point in time, especially if subjects adopt the strategy of executing the task serially, and hence effort may not have been effectively manipulated. A similar criticism applies to Craik and Tulving (1975, Experiment 5). It is apparent, then, that effort cannot be so lightly disregarded.

It is the purpose of this research to provide a more valid test of effort as a causal factor in word recall. To this end, four experiments were run. The first was designed to determine if experimenter-defined manipulations of effort within single cognitive tasks would indeed affect recall. The second attempted to establish

an independent measurement criterion for the effort continuum. The third and fourth experiments were designed to rule out possible alternative explanations of effort effects; in addition, the fourth employed a modification of the earlier designs, changed to remove certain anomalies observed in previous results.

Experiment 1

In this experiment, the standard levels-of-processing paradigm was employed. Subjects experienced a series of nonsemantic (anagram) and semantic (sentence-completion) tasks, after which they were surprised with a word-recall test. For each type of task, there were two different degrees of effort, the difference corresponding to a manipulation believed on logical grounds to correlate with momentary central-processing capacity demands. For the anagram task, subjects attempted to unscramble a series of letters to determine which of two words contained the same letters. In the low-effort condition, the anagrams were scrambled very little; in the high-effort condition, letter rearrangement was extensive.

For the sentence-completion task, subjects tried to choose which of two presented words fit into a blank in a given sentence. For low-effort sentences a single word was strongly implied, whereas for high-effort sentences the best-fitting word was not readily apparent.

The major hypothesis of this study was that words encountered during high-effort tasks would be better 'recalled than those in the low-effort condition. This was an obvious derivation for target words (correct choices), and was also extended to distractor words on the assumption that, being simultaneously presented with the target words, they could share in the processing effort required by the task.

Method

Materials. Eighty nouns of 5–8 letters were selected from Paivio, Yuille, and Madigan's (1968) norms and were of uniformly high imagery, meaningfulness, and frequency values. For each word, four

task slides containing two anagrams and two sentences were constructed. The anagrams differed in terms of the number of letter positions altered from those in the source word; this manipulation produced low-effort and high-effort anagrams. For example, the word *doctor* was transformed into the anagrams *dortoc* (low effort) and *croodt* (high effort). For the sentence slides, sentences were constructed containing a blank into which the word could meaningfully be inserted. The blanks occurred in the same position in each sentence, generally at the end, and served the same syntactic role in the two sentences. The sentences differed only in the degree to which each implied the missing word. Thus, for low-effort sentences the missing word was judged by the experimenters to be almost redundant with other information in the sentence: The girl was awakened by her frightening ____. For high-effort sentences the missing word was judged to be less determined by the information in the sentence: The man was alarmed by the frightening ____. For the above examples, the missing word was *dream*; it has a relatively high probability of being chosen for the low-effort sentence and is one of several possibilities for the high-effort sentence. For this task an attempt was also made to equalize the imagery value and salience of the two sentences.

Target slides were composed of the target word and a distractor, one appearing above the other, with relative position being counterbalanced so that target words occurred in the top position for half of the slides. Distractors were chosen to minimize choice errors in the sentence completion task. In general, completing the sentence with the distractor yielded a completely meaningless statement.

Subjects and design. Sixteen students from introductory psychology classes were assigned separate lists. Each list was composed of 40 target–distractor pairs. Ten pairs were presented within each task condition (low-effort anagram, high-effort anagram, low-effort sentence, high-effort sentence). Furthermore, target–distractor positions on the target slide were counterbalanced within condition, and list order was completely randomized. Therefore, each subject saw all 80 words in a unique order. Across subjects, each word was presented twice as a target (once above the distractor and once below) and twice as a distractor in each task condition.

Procedure. Subjects were run individually in an incidental-learning paradigm. They were seated before a rear projection screen, with two vertically arranged buttons situated on the table in front of them. The buttons were labeled *top* and *bottom*, and there was a resting place for the finger indicated between them. Subjects were told at the start of the session that they would see a series of slide pairs, each pair appearing simultaneously on the screen. When an anagram appeared on the task slide, the subjects' task was to determine to which of two words on the target slide the anagram corresponded, and to indicate their choice by pressing the corresponding button, top or bottom, with their preferred hand. When an incomplete sentence appeared on the task slide, the task was to choose
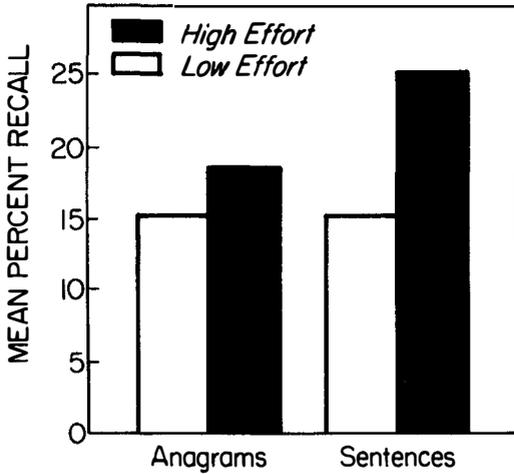
Figure 1. Mean percent recall of words chosen as correct responses in high- and low-effort anagrams and sentence tasks.

the word that made most sense in the sentence. Target slides were presented on the left side of the screen, and task slides on the right, both shown simultaneously. Subjects were informed that this was a reaction-time experiment, although they were to be sure of their choice before responding. Each trial was preceded by a warning light, and five practice trials were given. Finally, a constant interval of 9 sec was observed from the onset of one slide pair to the onset of the next.

Following the processing tasks, subjects engaged in 30 sec of simple arithmetic and were then asked to free recall all words appearing on the target slides. Guessing was encouraged by the instructions, and all subjects were required to guess at the end of their initial recall periods.

Results and Discussion

Overall recall. We first examined the mean percent word recall using an analysis of variance with three within-subject factors: word type (target vs. distractor), degree of effort, and level of processing. In this and all future analyses, values of $p < .05$ were considered significant. More words were recalled in the high-effort (18.6%) than in the low-effort condition (11.1%), $F(1, 15) = 8.50$, $MS_e = 16.93$. A secondary finding was that recall of correctly chosen words was greater than recall of distractor words, $F(1, 15) = 18.64$, $MS_e = 8.38$; however, this factor did not interact with either of the other two factors. Finally, there was no effect of level of

processing, words in the sentence-completion task being recalled only slightly better than words in the anagram task.

Target recall. Figure 1 presents the mean percent recall of target (correct) words in each of the four task conditions. A second analysis of variance for the effects of effort and level of processing was carried out for target words alone. Neither the main effects of effort nor levels of processing for target words were reliable, nor was there any Effort × Level of Processing interaction. A separate analysis of effort effects within each level of processing revealed that high-effort sentence tasks led to better recall than low-effort sentence task, $F(1, 15) = 4.91$, $MS_e = 2.86$, although there was no significant effect of effort for the anagram task.

Distractor recall. An analysis of recall for distractor words showed a pattern of results similar to recall of target words. For distractor (nontarget) words, the mean percent recall for low- and high-effort anagrams was 5.6% and 11.3%, respectively, and 7.5% and 19.4% for low- and high-effort sentences, respectively. The overall effect of effort on distractor recall was significant, $F(1, 15) = 10.65$, $MS_e = 4.60$, whereas there was neither a reliable effect of levels of processing nor a significant interaction of these two factors. Finally, within the sentence-completion task, effort had a reliable influence, $F(1, 15) = 6.45$, $MS_e = 3.49$; within the anagram task, the effort effect was not reliable.

Experiment 2

Experiment 1 provided substantial support for the role of cognitive effort in recall. The major purpose of the second experiment was to establish an independent measure to assess the effort required for a given task, and to use the measure to validate the effort manipulation employed in the first experiment. In addition, a slight modification of the original task was made in order to magnify the impact of the effort manipulation. It was felt that since the effort manipulation was confined to the relation between the target word and

the task, including a distractor with the target might have modified the effort actually required for a given task. Therefore, a change was made from the two-choice reaction-time task to a simple yes–no decision task.

The logic of this experiment essentially follows that presented by Kerr (1973) in her review of secondary task paradigms, and the specific design used is quite similar to that of Ellis and Kreezer (Note 1) and Johnston and Uhl (1976). The basic idea involves adding a secondary tone-detection task to the primary task, the assumption being that the greater the proportion of the limited-capacity central processor required for the primary task, the worse the performance will be on the secondary task. The primary task, as emphasized by instructions, was the same for this experiment as the task used in Experiment 1, with the exception that instead of choosing between two words, subjects simply determined whether a single word was correct (i.e., fit in the sentence or could be made from the scrambled letters). At the same time subjects performed the primary task, they were required to press another button as quickly as possible whenever they heard a tone over their headphones, and reaction times for tone detection were recorded.

Two major predictions were derived from effort theory. First, it was expected that correct words would in general be more frequently recalled if they appeared in the high-effort tasks than if they were present in low-effort tasks. No predictions were made regarding words for which a negative response was appropriate, since the effort manipulation bore no obvious relationship to the incorrect words. Second, measures of tone-detection reaction times were expected to reflect directly the effort needed for a given type of task, with high-effort tasks showing longer reaction times than low-effort tasks.

Method

Materials, subjects, and design. Task slides were identical to those in Experiment 1; however, target slides contained only one word that was either the correct word for the task or the incorrect
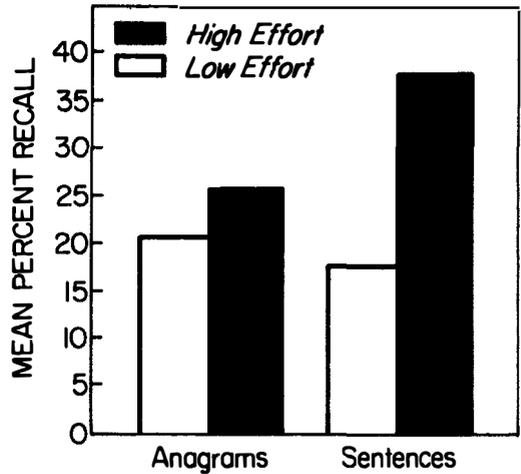

Figure 2. Mean percent recall of words for which positive responses were appropriate, separated according to level of processing and effort condition.

distractor previously paired with the word in Experiment 1. Sixteen subjects from introductory psychology classes received the same lists of tasks as did those in Experiment 1. The new corresponding targets were counterbalanced within task condition so that half would yield *yes* responses and half would yield *no* responses. Thus, each subject saw 40 words, and across subjects each word was presented once for a positive response and once for a negative response in each task condition.

Procedure. Subjects were run individually in an incidental learning paradigm that included a disjunctive reaction-time primary task and a tone-detection secondary task (probe task). They were seated at a table in front of a rear projection screen. On one side of the table were two buttons labeled *yes* and *no*, and on the other was a button labeled *tone*. These could be shifted according to the subjects' handedness. The two tasks were then described to the subjects. For the primary task, the subjects determined whether the target word on the left side of the screen could be used to complete the task on the right side of the screen, that is, whether it could be constructed from scrambled letters for anagram task slides or fitted in a sentence blank for sentence-completion task slides. For the secondary task, they were told that on some trials a tone would occur over the left earphone. Upon its occurrence, they were to press the tone button with their nonpreferred hand and continue with the primary task. It was emphasized that the primary task was the one of most importance. Probe onset occurred either 500, 1,000, 1,500, or 2,000 msec following the onset of the slide pair, or not at all. The probe was a weak 1000-Hz tone of approximately 14 dB (SPL). The five probe conditions were counterbalanced within task conditions and response (yes, no) conditions.

Table 1
*Mean Probe Reaction Times (in msec)
for Positive and Negative Responses*

Primary task condition	Primary task response	
	Positive	Negative
Low-effort anagram	1,155	1,047
High-effort anagram	1,205	1,140
Low-effort sentence	937	1,204
High-effort sentence	1,225	1,137

In all other respects, the procedure was identical to that of Experiment 1.

Results and Discussion

Percent recall. Figure 2 presents mean percent recall of words for which positive responses were appropriate, presented according to the type of processing task. As in the first experiment, an analysis of variance for these words showed that the overall effect of effort was significant, $F(1, 15) = 6.82$, $MS_e = 3.67$, and that there was no reliable levels-of-processing effect and no Effort $\times$ Level of Processing interaction. Further analysis of the effect of effort within each level of processing established a substantial effect within sentence-completion tasks, $F(1, 15) = 10.90$, $MS_e = 1.46$, but no reliable effect within anagram tasks.

Percent recall for negative responses was much lower than for positive responses (10.0% and 7.5% for low- and high-effort anagrams, and 23.8% and 27.5% for low- and high-effort sentences). As revealed by an analysis of variance, effort was not a reliable factor in accounting for negative-response recall, although there was an effect of level of processing for negative responses, $F(1, 15) = 15.73$, $MS_e = 2.89$.

Probe reaction time. Table 1 reports mean probe response latencies for both positive and negative responses in the primary processing task. The means represent response times for all four probe onset latencies and exclude response times to probes occurring after the response to the primary task was executed. (A preliminary analysis determined that there were no differential effects due to time of probe onset.) The most important finding was that effort differences in reaction time corresponded to effort differences in percent recall. An analysis of variance applied to probe latencies for positive primary-task responses established that such responses were longer during the high-effort processing tasks, $F(1, 15) = 5.41$, $MS_e = .40$. Differences in probe reaction time across levels of processing were not reliable, nor was the Effort $\times$ Level of Processing interaction. Within tasks, probe latencies were longer for the high-effort sentence-completion task than for the low-effort sentence task, $F(1, 15) = 15.00$, $MS_e = .09$, but there was no such effect for anagrams. Finally, no reliable effects of either effort or level of processing were obtained for negative responses in the processing task.

Several points deserve emphasis. First, the use of a secondary task to provide an independent measure of effort was shown to be successful. The pattern of significant effort effects for both positive and negative responses in recall performance was mirrored exactly in the probe reaction-time data, even extending to effects within each task. Also, it is noteworthy that in spite of a significant level-of-processing effect for the recall of negative response words, there was no such effect for probe reaction times. This clearly indicates that probe reaction-time performance does measure something separate from and independent of *depth of processing.*

The second point to be made is that the manipulations used to vary effort, based on a logical rather than an empirical approach, were generally successful and were enhanced by using a simple reaction-time task instead of a two-choice task. Apparently, the required central-processing capacity was indeed being manipulated.

The last point relates to the lack of a significant level-of-processing effect in the first experiment. It is possible that assuming the overriding importance of effort, this outcome might have been due to an absence of any difference in effort between these particular nonsemantic and semantic tasks. This speculation is supported by the

reaction-time data of this second study, which essentially show no overall difference between the two levels of processing. An alternative explanation will be considered in the fourth experiment.

Experiment 3

There is an alternative account of the findings of the preceding experiments demonstrating effort effects, namely, that the determining factor is not the difference due to variations in momentary effort, rather it is the total time required for each task that affects recall. High-effort tasks may simply take longer to perform than low-effort tasks, and this alone may lead to better recall. Unfortunately, more is needed to assess this possibility than simply looking at reaction times for the primary task. This is so because total reaction time includes the time required to read the task slide, and the latter undoubtedly differs for the different task conditions. Thus, it almost certainly takes longer for the average subject to read a sentence than to read a near word (i.e., anagram); similarly, an anagram in the high-effort condition, being less like a real word, may well take longer to read than one in the low-effort condition. The dependent variable needed to rule out the total time explanation is one that corrects total reaction times for differences in reading times. The purpose of this experiment was to obtain an estimate of this dependent variable by separately assessing the total reaction time and the task reading time for the same subject on each task, and then analyzing the differences between task conditions using the resultant adjusted dependent variable.

Method

Materials, subjects, and design. Sixteen subjects from introductory psychology courses participated, each being assigned a different list of task and target items. Specific tasks and targets, as well as specific orders, were identical to those used in Experiment 2.

Procedure. The experiment consisted of two phases, the first to assess task reaction times and the second to evaluate task reading times. The two phases were conducted 2 days apart in order to minimize possible transfer effects, and subjects were run individually. The first phase consisted of a series of yes–no decision tasks identical to those used in the incidental learning phase of Experiment 2. The instructions and general setup of equipment were also identical. Following the first phase, each subject was instructed to return 2 days later for another session and was then dismissed. In the second phase, the subjects were told that they would now see the same task slides as had been seen in the first session, but without any associated target slides. The task was simply to read each task slide, then press the top button used for yes responses as soon as the reading of the slide was completed. Each subject was then shown the 40 task slides, and reading times were recorded.

Results and Discussion

A new dependent variable, decision time, was computed by subtracting the reading time for each task from the corresponding reaction time to perform that task with a given target item. This was done separately for each subject, thereby producing a new set of 40 decision times for each. Decision time thus reflected the time needed to determine the appropriate response for a given task–target pairing, with the time required for reading the task removed.

The mean decision times for correct yes decisions were as follows: for low-effort anagrams, 2,018 msec; for low-effort sentences, 1,287 msec; for high-effort anagrams, 1,758 msec; and for high-effort sentences, 1,563 msec. An analysis of variance verified that the effect of effort was nonsignificant, although there was a level-of-processing effect, anagram tasks requiring longer to perform than sentence-completion tasks, $F(1, 15) = 14.17$. $MS_e = .96$. There was no interaction of these two factors. Though uninformative, a similar analysis of correct no decisions showed an absence of any reliable effects of level of processing or of effort.

It appears, then, that effort effects cannot be readily accounted for by an appeal to the total time hypothesis. When reaction times are corrected for reading time variations, there is virtually no difference between low-effort and high-effort tasks.

Experiment 4

There were two main reasons for conducting the final experiment. The first was to demonstrate further the inadequacy of the total-time explanation of the obtained effort effects, and the second was to remove two anomalies encountered in previous experiments.

The first goal was achieved by changing the nature of the task from one in which both target and task slide were presented simultaneously and remained present throughout task execution, to one in which the target slide was presented only briefly and appeared prior to the task slide. By making this change, any possibility that effort effects were arising due to greater time spent by subjects on the physically present target slide during high-effort tasks could be eliminated.

To explain how the second goal was achieved, the two anomalies observed in the earlier experiments need to be specified. One anomaly centers on the consistent failure to obtain a reliable effort effect for the anagram task alone, despite the fact that this was obtained for the sentence-completion task in both of the first two experiments and despite the presence in the second experiment of an overall effect of effort. A second anomaly concerns the failure in each of the preceding studies to obtain a reliable level-of-processing effect for correct positive responses.

The first anomaly may have been due to the fact that the target item was physically present throughout the task. Thus, in the anagram task, subjects may simply have scanned back and forth between target and task slides directly comparing letters, and the effect of differences in degree of letter rearrangement may thereby have been much reduced. It was hoped, therefore, that the previously mentioned change in this experiment from simultaneous to successive presentation of target–task slide pairs would result in the attainment of an effort effect for anagrams.

The second anomaly, lack of level-of-processing effects, may have been due to the intermixing of anagram and sentence-completion tasks. Given that the nature of the task was never known in advance and that subjects were instructed to read the target slide first, all words may have been processed to a fairly deep level, whether appearing in an anagram or a sentence-completion task. To reduce this difficulty, the paradigm was further modified. Sentence-completion and anagram tasks were blocked so that a number of tasks of the same type occurred for a series of trials, and subjects were informed in advance of the type of task that would occur within each given block. It was anticipated that as a result of these changes, subjects would be more inclined to process the target item only to the depth required by the task.

Method

Materials, subjects, and design. Subjects were 32 volunteers from introductory psychology courses. The specific target and task slides used were essentially the same as those employed in Experiment 2. However, slides were arranged for a given subject so that there was a series of sentence-completion tasks followed by a series of anagram tasks, and vice-versa, composing a total of four blocks. In addition, five trials at the beginning of the experiment were treated as practice trials, as was the first trial of each new block, and target items that were recalled later from these trials were not scored. Furthermore, a color slide was used to signal the nature of the task to be performed within the upcoming block of trials: A blue slide meant the tasks would all involve anagrams, a red slide signaled the sentence-completion task, and a yellow slide at the end of the entire series of slides meant that the primary task was no longer in effect and only probe detection would be necessary for the remainder of the trials.

For half of the subjects, the first and third blocks consisted of anagram tasks, the second and fourth of sentence-completion tasks; for the other half, this order was reversed. Across subjects, each word appeared in each block in each task condition, and in each of these conditions the response of yes was appropriate for half of the subjects and the response of no for the other half. Each subject, therefore, experienced a unique ordering of task and target slides. There was a total of 40 slide pairs, excluding practice slides, shown to each subject. Thus a typical series of slides might be blue slide, 15 anagram tasks (including five practice trials); red slide, 11 sentence-completion tasks (including one practice trial); blue slide, 11 anagram tasks; red slide, 11 sentence-completion tasks; yellow slide.

Procedure. The equipment setup was the same

as in Experiment 2, as were the primary and secondary tasks. Subjects were informed of the nature of the primary task and were further instructed that a given type of task would occur for a number of trials, the nature of that task being signaled by a preceding color slide. Which color signaled which task was fully explained until subjects said they clearly understood the relationship. Subjects were then told about the tone-detection task and were instructed to press the appropriate button whenever a tone was heard over the headphones. It was stressed that they should do this as quickly as possible, although it was most important that they do well on the primary task. Subjects were then shown all target–task slide pairs. The tone, as in Experiment 2, occurred either 500, 1,000, 1,500, or 2,000 msec after task slide onset, or not at all; for each type of task, there was an equal number of probes at each onset time. Following the appearance of the yellow slide at the end of the entire series of slide pairs, five trials of tone detection alone were administered in order to reduce recency effects in recall. Finally, subjects were asked to recall as many of the words seen during the primary task as they could remember. Guessing was encouraged, and subjects were required to produce at least 20 words.

Results and Discussion

Percent recall. Figure 3 shows the mean percent recall of words for which positive responses were appropriate, divided according to specific task condition. For these words, an analysis of variance revealed main effects of both effort, $F(1, 31) = 28.97$, $MS_e = 2.18$, and level of pro-

Figure 3. Mean percent recall of words for which positive responses were appropriate, separated by level of processing and effort condition.

Table 2
Mean Probe Reaction Times (in msec) for Positive and Negative Responses

Primary task condition	Primary task response	
	Positive	Negative
Low-effort anagram	745	747
High-effort anagram	953	749
Low-effort sentence	791	831
High-effort sentence	1,127	856

cessing, $F(1, 31) = 5.24$, $MS_e = 3.73$, with no interaction of these two factors. A further analysis of effort effects within each level of processing established reliable effects for both sentences, $F(1, 31) = 9.62$, $MS_e = 1.87$, and anagrams, $F(1, 31) = 9.89$, $MS_e = 1.39$.

Words for which negative responses were appropriate only showed a significant level of processing effect, $F(1, 31) = 11.33$, $MS_e = 3.38$. There was no effort effect and no reliable interaction for these words.

Probe reaction time. All probes that appeared after the subject responded to the primary task were eliminated from the analysis. As in Experiment 2, there was no effect of time of probe onset, so all probe reaction times were combined within each task condition. Separate analyses of variance were performed for positive and negative responses. Table 2 presents the mean probe response latencies for both positive and negative responses, divided according to primary task condition. For positive responses, probe latencies were longer in the high-effort than in the low-effort primary tasks, $F(1, 31) = 19.79$, $MS_e = .48$. There were no differences in probe latencies for different levels of processing, nor was there any interaction. Within each level of processing, effort had a reliable influence: for anagrams, $F(1, 31) = 18.83$, $MS_e = .07$; for sentences, $F(1, 31) = 11.34$, $MS_e = .32$. For negative responses there were no reliable effects of effort or level of processing, nor was there an interaction of these factors.

Thus, the adjustments made in this experiment did eliminate the anomalies observed in the earlier studies. Word recall
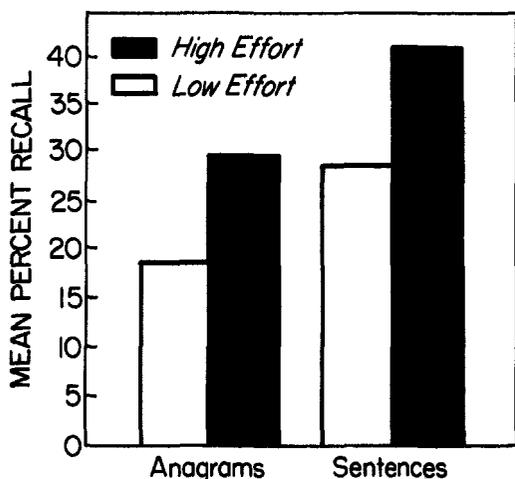
for items appearing in positive response tasks was not only higher for high-effort tasks generally, but was also reliably higher for high-effort anagrams than for low-effort anagrams and for high-effort sentences over low-effort sentences when these effects were looked at separately. A substantial level-of-processing effect was also obtained for positive as well as negative response items in this experiment.

Furthermore, the adequacy of probe response times as an index of cognitive effort once more received substantial justification. Not only were probe response latencies for positive response items longer in the high-effort conditions taken as a whole, but as in word recall, they were also longer for high-effort anagrams and for high-effort sentences when separately compared to their low-effort counterparts. Also, there were no probe reaction-time differences between anagram and sentence-completion tasks for either negative or positive responses, despite the fact that there were level of processing effects evident in word recall for both types of response. This once again points to the independence of cognitive effort from the depth of processing notion.

General Discussion

The major conclusions from this research are threefold: (a) Effort can be orthogonally varied within different levels of processing; (b) the amount of effort required by a task is an important determinant of later recall performance, with greater effort leading to greater recall; (c) there is available an independent, sensitive measurement criterion of effort, namely, secondary task performance, and this metric measures something separate from the level of processing involved. The viability of the cognitive effort concept as an explanatory factor in memory research, as well as its relative advantage over the levels-of-processing view, given its susceptibility to precise quantitative specification, has thus been supported.

Several questions warranting further investigation are suggested by these conclu-

sions, only two of which are mentioned. One concerns determination of which manipulations of task requirements and specific stimulus attributes correspond to manipulations of effort. This determination can, of course, be simply made through a strict empirical approach, that is, by evaluating the effects of various manipulations on secondary task performance, assuming a secondary task of sufficient sensitivity. A taxonomy of those stimulus and task dimensions closely tied to the effort continuum would potentially be of considerable theoretical interest, since such dimensions would presumably correspond to the factors of importance in predicting memorial performance.

A second question pertains to the mechanism whereby differences in effort lead to differences in recall. It could be that with greater effort, there is a greater tendency to integrate the context (or task environment) and the target word, thereby facilitating later retrieval, or it might be the case that an item in a higher effort situation is stored in memory as a trace of greater strength. A number of techniques could be used to assess these various possibilities. These might include the use of memory paradigms other than free recall to determine the relative importance of retrieval operations and the addition of an intentionality dimension, that is, varying whether subjects are informed of the final memory test, to evaluate the extent to which effort is under voluntary control.

Certainly, the experiments reported here raise numerous questions deserving further investigative consideration. It is hoped that the cognitive effort concept will provide a framework for such investigations that lacks some of the shortcomings evident in the depth-of-processing approach.

Reference Notes

1. Ellis, H. C., & Kreezer, G. L. *Modification of the audibility curve by attention to a distraction task and its treatment by transfer-function concepts.* Paper presented at the Midwestern Psychological Association, St. Louis, May 1956.
2. Ellis, H. C. *Strategies in human memory.* Presidential address to the Southwestern Psychological Association, New Orleans, April 1978.

References

Baddeley, A. D. The trouble with levels: A re-examination of Craik and Lockhart's framework for memory research. *Psychological Review*, 1978, *85*, 139–152.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 671–684.

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.

Ellis, H. C., Parente, F. J., Grah, C. R., & Spiering, K. Coding strategies, perceptual grouping, and the "variability effect" in free recall. *Memory & Cognition*, 1975, *3*, 226–232.

Ellis, H. C., Parente, F. J., & Walker, C. W. Coding and varied input versus repetition in human memory. *Journal of Experimental Psychology*, 1974, *102*, 284–290.

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.

Johnston, W. A., & Uhl, C. N. The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 153–160.

Johnston, W. A., Wagstaff, R. R., & Griffith, D. Information-processing analysis of verbal learning. *Journal of Experimental Psychology*, 1972, *96*, 307–314.

Kahneman, D. *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Keele, S. W. *Attention and human performance,* Pacific Palisades, Calif.: Goodyear, 1973.

Kerr, B. Processing demands during mental operations. *Memory & Cognition*, 1973, *1*, 401–412.

Moray, N. Where is capacity limited? A survey and a model. In A. Sanders (Ed.), *Attention and performance.* Amsterdam: North-Holland, 1967.

Nelson, D. L., Walling, J. R., & McEvoy, C. L. Doubts about depth. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, *5*, 24–44.

Nelson, T. O. Repetition and depth of processing. *Journal of Verbal and Verbal Behavior*, 1977, *16*, 151–171.

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 concrete nouns. *Journal of Experimental Psychology Monograph*, 1968, *76*(1, Pt. 2).

Posner, M. I., & Snyder, C. R. R. Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium.* Hillsdale, N.J.: Erlbaum, 1975.

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

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.