Journal of Experimental Psychology: General
1975, Vol. 104, No. 3, 268–294

Depth of Processing and the Retention of Words in Episodic Memory

Fergus I. M. Craik and Endel Tulving
University of Toronto, Toronto, Ontario, Canada

SUMMARY

Ten experiments were designed to explore the levels of processing framework for human memory research proposed by Craik and Lockhart (1972). The basic notions are that the episodic memory trace may be thought of as a rather automatic by-product of operations carried out by the cognitive system and that the durability of the trace is a positive function of "depth" of processing, where depth refers to greater degrees of semantic involvement. Subjects were induced to process words to different depths by answering various questions about the words. For example, shallow encodings were achieved by asking questions about typescript; intermediate levels of encoding were accomplished by asking questions about rhymes; deep levels were induced by asking whether the word would fit into a given category or sentence frame. After the encoding phase was completed, subjects were unexpectedly given a recall or recognition test for the words. In general, deeper encodings took longer to accomplish and were associated with higher levels of performance on the subsequent memory test. Also, questions leading to positive responses were associated with higher retention levels than questions leading to negative responses, at least at deeper levels of encoding.

Further experiments examined this pattern of effects in greater analytic detail. It was established that the original results did not simply reflect differential encoding times; an experiment was designed in which a complex but shallow task took longer to carry out but yielded lower levels of recognition than an easy, deeper task. Other studies explored reasons for the superior retention of words associated with positive responses on the initial task. Negative responses were remembered as well as positive responses when the questions led to an equally elaborate encoding in the two cases. The idea that elaboration or "spread" of encoding provides a better description of the results was given a further boost by the finding of the typical pattern of results under intentional learning conditions, and where each word was exposed for 6 sec in the initial phase. While spread and elaboration may indeed be better descriptive terms for the present findings, retention depends critically on the qualitative nature of the encoding operations performed; a minimal semantic analysis is more beneficial than an extensive structural analysis.

Finally, Schulman's (1974) principle of congruity appears necessary for a complete description of the effects obtained. Memory performance is enhanced to the extent that the context, or encoding question, forms an integrated unit with the word presented. A congruous encoding yields superior memory performance because a more elaborate trace is laid down and because in such cases the structure of semantic memory can be utilized more effectively to facilitate retrieval. The article concludes with a discussion of the broader implications of these data and ideas for the study of human learning and memory.

While information-processing models of human memory have been concerned largely with structural aspects of the system, there is a growing tendency for theorists to focus, rather, on the *processes* involved in learning and remembering. Thus the theorist's task, until recently, has been to provide an adequate description of the characteristics and interrelations of the successive stages through which information flows. An alternative approach is to study more directly those processes involved in remembering—processes such as attention, encoding, rehearsal, and retrieval—and to formulate a description of the memory system in terms of these constituent operations. This alternative viewpoint has been advocated by Cermak (1972), Craik and Lockhart (1972), Hyde and Jenkins (1969, 1973), Kolers (1973a), Neisser (1967), and Paivio (1971), among others, and it represents a sufficiently different set of fundamental assumptions to justify its description as a new paradigm, or at least a miniparadigm, in memory research. How should we conceptualize learning and retrieval operations in these terms? What changes in the system underlie remembering? Is the "memory trace" best regarded as some copy of the item in a memory store (Waugh & Norman, 1965), as a bundle of features (Bower, 1967), as the record resulting from the perceptual and cognitive analyses carried out on the stimulus (Craik & Lockhart, 1972), or do we remember in terms of the encoding operations themselves (Neisser, 1967; Kolers, 1973a)? Although we are still some way from answering these crucial questions satisfactorily, several recent studies have provided important clues.

The incidental learning situation, in which subjects perform different orienting tasks, provides an experimental setting for the study of mental operations and their effects on learning. It has been shown that when subjects perform orienting tasks requiring analysis of the meaning of words in a list, subsequent recall is as extensive and as highly structured as the recall observed under intentional conditions in the absence of any specific orienting task; further research has indicated that a "process" explanation is most compatible with the results (Hyde, 1973; Hyde & Jenkins, 1969, 1973; Walsh & Jenkins, 1973). Schulman (1971) has also shown that a semantic orienting task is followed by higher retention of words than a "structural" task in which the nonsemantic aspects of the words are attended to. Similar findings have been reported for the retention of sentences (Bobrow & Bower, 1969; Rosenberg & Schiller, 1971; Treisman & Tuxworth, 1974) and in memory for faces (Bower & Karlin, 1974). In all these experiments, an orienting task requiring semantic or affective judgments led to better memory performance than tasks involving structural or syntactic judgments. However, the involvement of semantic analyses is not the whole story: Schulman (1974) has shown that congruous queries about words (e.g., "Is a SOPRANO a singer?") yield better memory for the words than incongruous queries (e.g., "Is MUSTARD concave?"). Instruction to form images from the words also leads to excellent retention (e.g., Paivio, 1971; Sheehan, 1971).

The results of these studies have important theoretical implications. First, they demonstrate a continuity between incidental and intentional learning—the operations carried out on the material, not the intention to learn, as such, determine retention. The results thus corroborate Postman's (1964) position on the essential similarity of incidental and intentional learning, although the recent work is more usually described in terms of similar processes rather than similar responses (Hyde & Jenkins, 1973). Second, it seems clear that attention to the word's meaning is a necessary prerequisite of good retention. Third, since retrieval

The research reported in this article was supported by National Research Council of Canada Grants A8261 and A8632 to the first and second authors, respectively. The authors gratefully acknowledge the assistance of Michael Anderson, Ed Darte, Gregory Mazuryk, Marsha Carnat, Marilyn Tiller, and Margaret Barr.

Requests for reprints should be sent to F. I. M. Craik, Erindale College, University of Toronto, Mississauga, Ontario, L5L 1C6, Canada.

conditions are typically held constant in the experiments described above, the differences in retention reflect the effects of different encoding operations, although the picture is complicated by the finding that different encoding operations are optimal for different retrieval conditions (e.g., Eagle & Leiter, 1964; Jacoby, 1973). Fourth, large differences in recall under different encoding operations have been observed under conditions where the subjects' task does not entail organization or establishment of interitem associations; thus the results seem to take us beyond associative and organization processes, as important determinants of learning and retention. It may be, of course, that the orienting tasks actually do lead to organization as suggested by the results of Hyde and Jenkins (1973). Yet, it now becomes possible to entertain the hypothesis that optimal processing of individual words, qua individual words, is sufficient to support good recall. Finally, the experiments may yield some insights into the nature of learning operations themselves. Classical verbal learning theory has not been much concerned with processes and changes within the system but has concentrated largely on manipulations of the material or the experimental situation and the resulting effects on learning. Thus at the moment, we know a lot about the effects of meaningfulness, word frequency, rate of presentation, various learning instructions, and the like, but rather little about the nature and characteristics of underlying or accompanying mental events. Experimental and theoretical analysis of the effects of various encoding operations holds out the promise that intentional learning can be reduced to, and understood in terms of, some combination of more basic operations.

The experiments reported in the present paper were carried out to gain further insights into the processes involved in good memory performance. The initial experiments were designed to gather evidence for the depth of processing view of memory outlined by Craik and Lockhart (1972). These authors proposed that the memory trace could usefully be regarded as the byproduct of perceptual processing; just as perception may be thought to be composed of a series of analyses, proceeding from early sensory processing to later semantic-associative operations, so the resultant memory trace may be more or less elaborate depending on the number and qualitative nature of the perceptual analyses carried out on the stimulus. It was further suggested that the durability of the memory trace is a function of depth of processing. That is, stimuli which do not receive full attention, and are analyzed only to a shallow sensory level, give rise to very transient memory traces. On the other hand, stimuli that are attended to, fully analyzed, and enriched by associations or images yield a deeper encoding of the event, and a long-lasting trace.

The Craik and Lockhart formulation provides one possible framework to accommodate the findings from the incidental learning studies cited above. It has the advantage of focusing attention on the processes underlying trace formation and on the importance of encoding operations; also, since memory traces are not seen as residing in one of several stores, the depth of processing approach eliminates the necessity to document the capacity of postulated stores, to define *the* coding characteristic of each store, or to characterize the mechanism by which an item is transferred from one store to another. Despite these advantages, there are several obvious shortcomings of the Craik and Lockhart viewpoint. Does the levels of processing framework say any more than "meaningful events are well remembered"? If not, it is simply a collection of old ideas in a somewhat different setting. Further, the position may actually represent a backward step in the study of human memory since the notions are much vaguer than any of the mathematical models proposed, for example, in Norman's (1970) collection. If we already know that the memory trace can be precisely represented as

$$l = \lambda e^{-\psi t^{1-\gamma}}$$

(Wickelgren, 1973), then such woolly statements as "deeper processing yields a more durable trace" are surely far behind us. Third, and most serious perhaps, the very least the levels position requires is some independent index of depth—there are obvious dangers of circularity present in that any well-remembered event can too easily be labeled *deeply processed.*

Such criticisms can be partially countered. First, cogent arguments can be marshaled (e.g., Broadbent, 1961) for the advantages of working with a rather general theory—provided the theory is still capable of generating predictions which are distinguishable from the predictions of other theories. From this general and undoubtedly true starting point, the concepts can be refined in the light of experimental results suggested by the theoretical framework. In this sense the levels of processing viewpoint will encourage rather different types of question and may yield new insights. A further point on the issue of general versus specific theories is that while strength theories of memory are commendably specific and sophisticated mathematically, the sophistication may be out of place if the basic premises are of limited generality or even wrong. It is now established, for example, that the trace of an event can be readily retrieved in one environment of retrieval cues, while it is retrieved with difficulty in another (e.g., Tulving & Thomson, 1973); it is hard to reconcile such a finding with the view that the probability of retrieval depends only on some unidimensional strength.

With regard to an independent index of processing depth, Craik and Lockhart (1972) suggested that, when other things are held constant, deeper levels of processing would require longer processing times. Processing time cannot always be taken as an absolute indicator of depth, however, since highly familiar stimuli (e.g., simple phrases or pictures) can be rapidly analyzed to a complex meaningful level. But within one class of materials, or better, with one specific stimulus, deeper processing is assumed to require more time. Thus, in the present studies, the time to make decisions at different levels of analysis was taken as an initial index of processing depth.

The purpose of this article is to describe 10 experiments carried out within the levels of processing framework. The first experiments examined the plausibility of the basic notions and attempted to rule out alternative explanations of the results. Further experiments were carried out in an attempt to achieve a better characterization of depth of processing and how it is that deeper semantic analysis yields superior memory performance. Finally, the implications of the results for an understanding of learning operations are examined, and the adequacy of the depth of processing metaphor questioned.

Experimental Investigations

Since one basic paradigm is used throughout the series of studies, the method will be described in detail at this point. Variations in the general method will be indicated as each study is described.

General Method

Typically, subjects were tested individually. They were informed that the experiment concerned perception and speed of reaction. On each trial a different word (usually a common noun) was exposed in a tachistoscope for 200 msec. Before the word was exposed, the subject was asked a question about the word. The purpose of the question was to induce the subject to process the word to one of several levels of analysis, thus the questions were chosen to necessitate processing either to a relatively shallow level (e.g., questions about the word's physical appearance) or to a relatively deep level (e.g., questions about the word's meaning). In some experiments, the subject read the question on a card; in others, the question was read to him. After reading or hearing the question, the subject looked in the tachistoscope with one hand resting on a *yes* response key and the other on a *no* response key. One second after a warning "ready" signal the word appeared and the subject recorded his (or her) decision by pressing the appropriate key (e.g., if the question was "Is the word an animal name?" and the word presented was TIGER, the subject would respond *yes*). After a series of such question and answer trials, the subject was unexpectedly given a retention test for the words. The expectation was that memory performance would vary systematically with the depth of processing.

Three types of question were asked in the initial encoding phase. (a) An analysis of the physical structure of the word was effected by asking about the physical structure of the word

TABLE 1

TYPICAL QUESTIONS AND RESPONSES USED IN THE EXPERIMENTS

Level of processing	Question	Answer Yes	Answer No
Structural	Is the word in capital letters?	TABLE	table
Phonemic	Does the word rhyme with WEIGHT?	crate	MARKET
Category	Is the word a type of fish?	SHARK	heaven
Sentence	Would the word fit the sentence: "He met a ____ in the street"?	FRIEND	cloud

(e.g., "Is the word printed in capital letters?"). (b) A phonemic level of analysis was induced by asking about the word's rhyming characteristics (e.g., "Does the word rhyme with TRAIN?"). (c) A semantic analysis was activated by asking either categorical questions (e.g., "Is the word an animal name?") or "sentence" questions (e.g., "Would the word fit the following sentence: 'The girl placed the ______ on the table'?"). Further examples are shown in Table 1. At each of the three levels of analysis, half of the questions yielded *yes* responses and half *no* responses.

The general procedure thus consisted of explaining the perceptual-reaction time task to a single subject, giving him a long series of trials in which both the type of question and *yes-no* decisions were randomized, and finally giving him an unexpected retention test. This test was either free recall ("Recall all the words you have seen in the perceptual task, in any order"); cued recall, in which some aspect of each word event was re-presented as a cue; or recognition, where copies of the original words were re-presented along with a number of distractors. In the initial encoding phase, response latencies were in fact recorded: A millisecond stop clock was started by the timing mechanism which activated the tachistoscope, and the clock was stopped by the subject's key response. Typically, over a group of subjects, the same pool of words was used, but each word was rotated through the various level and response combinations (CAPITALS?-*yes;* SENTENCE?-no, and so on). The general prediction was that deeper level questions would take longer to answer but would yield a more elaborate memory trace which in turn would support higher recognition and recall performance.

Experiment 1

Method. In the first experiment, single subjects were given the perceptual-reaction time test; this encoding phase was followed by a recognition test. Five types of question were used. First, "Is there a word present?" Second, "Is the word in capital letters?" Third, "Does the word rhyme with ______?" Fourth, "Is the word in the category ______?" Fifth, "Would the word fit in the sentence ______?" When the first type of question was asked ("Is there a word present?"), on half of the trials a word was present and on half of the trials no word was present on the tachistoscope card; thus, the subject could respond *yes* when he detected any wordlike pattern on the card. (This task may be rather different from the others and was not used in further experiments; also, of course, it yields difficulties of analysis since no word is presented on the negative trials, these trials cannot be included in the measurement of retention.)

The stimuli used were common two-syllable nouns of 5, 6, or 7 letters. Forty trials were given; 4 words represented each of the 10 conditions (5 levels × *yes-no*). The same pool of 40 words was used for all 20 subjects, but each word was rotated through the 10 conditions so that, for different subjects, a word was presented as a rhyme-*yes* stimulus, a category-*no* stimulus and so on. This procedure yielded 10 combinations of questions and words; 2 subjects received each combination. On each trial, the question was read to the subject who was already looking in the tachistoscope. After 2 sec, the word was exposed and the subject responded by saying *yes* or *no*—his vocal response activated a voice key which stopped a millisecond timer. The experimenter recorded the response latency, changed the word in the tachistoscope, and read the next question; trials thus occurred approximately every 10 sec.

After a brief rest, the subject was given a sheet with the 40 original words plus 40 similar distractors typed on it. Any one subject had actually only seen 36 words as no word was presented on negative "Word present?" trials. He was asked to check all words he had seen on the tachistoscope. No time limit was imposed for this task. Two different randomizations of the 80 recognition words were typed; one randomization was given to each member of the pair of subjects who received identical study lists. Thus each subject received a unique presentation-recognition combination. The 20 subjects were college students of both sexes paid for their services.

Results and discussion. The results are shown in Table 2. The upper portion shows response latencies for the different questions. Only correct answers were included in the analysis. The median latency was calculated for each subject; Table 2 shows mean medians. Although the five question levels were selected intuitively, the table shows that in fact response latency rises systematically as the questions necessitated deeper processing. Apart from the sentence level, *yes* and *no* responses took equivalent times. The median latency scores were subjected to an analysis of variance (after log transformation). The analysis showed a significant effect of level, $F(4, 171) = 35.4$, $p < .001$, but no effect of response type (*yes–no*) and no interaction. Thus, intuitively deeper questions —semantic as opposed to structural decisions about the word—required slightly longer processing times (150–200 msec).

Table 2 also shows the recognition results. Performance (the hit rate) increased substantially from below 20% recognized for questions concerning structural characteristics, to 96% correct for sentence–*yes* decisions. The other prominent feature of the recognition results is that the *yes* responses to words in the initial perceptual phase were accompanied by higher subsequent recognition than the *no* responses. Further, the superiority of recognition of *yes* words increased with depth (until the trend was apparently halted by a ceiling effect). These observations were confirmed by analysis of variance on recognition proportions (after arc sine transformation). Since the first level (word present?) had only *yes* responses, words from this level were not included in the analysis. Type of question was a significant factor, $F(3, 133) = 52.8$, $p < .001$, as was response type (yes–no), $F(1, 133) = 40.2$, $p < .001$. The Question × Response Type interaction was also significant, $F(3, 133) = 6.77$, $p < .001$.

The results have thus shown that different encoding questions led to different response latencies; questions about the surface form of the word were answered comparatively rapidly, while more abstract questions about the word's meaning took longer to answer. If processing time is an index of depth, then words presented after a semantic question were indeed processed more deeply. Further, the different encoding questions were associated with marked differences in recognition performance: Semantic questions were followed by higher recognition of the word. In fact, Table 2 shows that initial response latency is systematically related to subsequent recognition. Thus, within the limits of the present assumptions, it may be concluded that deeper processing yields superior retention.

TABLE 2

INITIAL DECISION LATENCY AND RECOGNITION PERFORMANCE FOR WORDS AS A FUNCTION OF INITIAL TASK (EXPERIMENT 1)

Response type	Level of processing 1	2	3	4	5
	Response latency (msec)				
Yes	591	614	689	711	746
No	590	625	678	716	832
	Proportion recognized				
Yes	.22	.18	.78	.93	.96
No		.14	.36	.63	.83

It is of course possible to argue that the higher recognition levels are more simply attributable to longer study times. This point will be dealt with later in the paper, but for the present it may be noted that in these terms, 200 msec of extra study time led to a 400% improvement in retention. It seems more reasonable to attribute the enhanced performance to qualitative differences in processing and to conclude that manipulation of levels of processing at the time of input is an extremely powerful determinant of retention of word events. The reason for the superior recognition of *yes* responses is not immediately apparent—it cannot be greater depth of processing in the simple sense, since *yes* and *no* responses took the same time for each encoding question. Further discussion of this point is deferred until more experiments are described.

Experiment 2 is basically a replication of Experiment 1 but with a somewhat tidier design and with more recognition distractors to remove ceiling effects.

Experiment 2

Method. Only three levels of encoding were used in this study: questions concerning type-

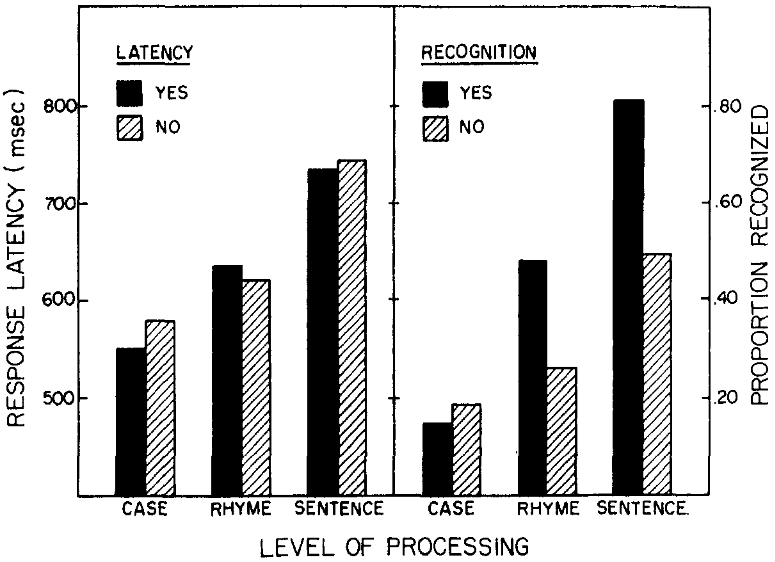

FIGURE 1. Initial decision latency and recognition performance for words as a function of the initial task (Experiment 2).

script (uppercase or lowercase), rhyme questions, and sentence questions (in which subjects were given a sentence frame with one word missing). During the initial perceptual phase 60 questions were presented: 10 *yes* and 10 *no* questions at each of the three levels. Question type was randomized within the block of 60 trials. The question was presented auditorily to the subject; 2 sec later the word appeared in the tachistoscope for 200 msec. The subject responded as rapidly as possible by pressing one of two response keys. After completing the 60 initial trials, the subject was given a typed list of 180 words comprising the 60 original words plus 120 distractors. He was told to check all words he had seen in the first phase.

All words used were five-letter common concrete nouns. From the pool of 60 words, two question formats were constructed by randomly allocating each word to a question type until all 10 words for each question type were filled. In addition, two orders of question presentation and two random orderings of the 180-word recognition list were used. Three subjects were tested on each of the eight combinations thus generated. The 24 subjects were students of both sexes paid for their services and tested individually.

Results and discussion. The left-hand panel of Figure 1 shows that response latency rose systematically for both response types, from case questions to rhyme questions to sentence questions. These data again are interpreted as showing that deeper processing took longer to accomplish. At each level, positive and negative responses took the same time. An analysis of variance on mean medians yielded an effect of question type, $F(2, 46) = 46.5$, $p < .001$, but yielded no effect of response type and no interaction.

Figure 1 also shows the recognition results. For *yes* words, performance increased from 15% for case decisions to 81% for sentence decisions—more than a fivefold increase in hit rate for memory performance for the same subjects in the same experiment. Recognition of *no* words also increased, but less sharply from 19% (case) to 49% (sentence). An analysis of variance showed a question type (level of processing) effect, $F(2, 46) = 118$, $p < .001$, a response type (*yes–no*) effect, $F(1, 23) = 47.9$, $p < .001$, and a Question Type × Response Type interaction, $F(2, 46) = 22.5$, $p < .001$.

Experiment 2 thus replicated the results of Experiment 1 and showed clearly (a) Different encoding questions are associated with different response latencies—this finding is interpreted to mean that semantic questions induce a deeper level of analysis of the presented word, (b) positive and negative responses are equally fast, (c) recognition increases to the extent that the encoding question deals with more abstract, semantic features of the word, and (d) words given a positive response are associated with higher recognition performance, but only after rhyme and category questions.

The data from Figure 1 are replotted in Figure 2, in which recognition performance is shown as a function of initial categorization time. Both *yes* and *no* functions are strikingly linear, with a steeper slope for *yes* responses. This pattern of data suggests that memory performance may simply be a function of processing *time* as such (regardless of "level of analysis"). This suggestion is examined (and rejected) in this article, where we argue that level of analysis, not processing time, is the critical determinant of recognition performance.

Experiments 3 and 4 extended the generality of these findings by showing that the same pattern of results holds in recall and under intentional learning conditions.

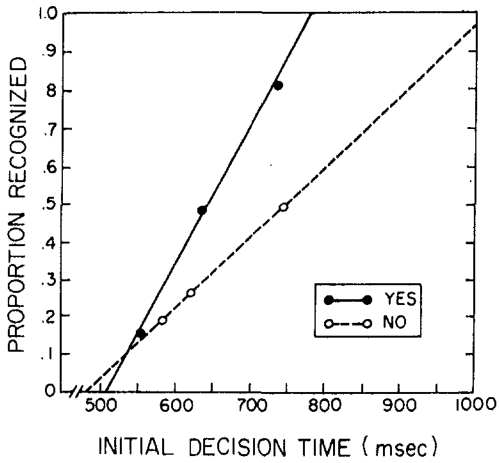

FIGURE 2. Proportion of words recognized as a function of initial decision time (Experiment 2).

Experiment 3

Method. Three levels of encoding were again included in the study by asking questions about typescript (case), rhyme, and sentences. On each trial the question was read to the subject; after 2 sec the word was exposed for 200 msec on the tachistoscope. The subject responded by pressing the relevant response key. At the end of the encoding trials, the subject was allowed to rest for 1 min and was then asked to recall as many words as he could. In Experiment 3, this final recall task was unexpected—thus the initial encoding phase may be considered an incidental learning task—while in Experiment 4, subjects were informed at the beginning of the session that they would be required to recall the words.

Pilot studies had shown that the recall level in this situation tends to be low. Thus, to boost recall, and to examine the effects of encoding level on recall more clearly, half of the words in the present study were presented twice. In all, 48 different words were used, but 24 were presented twice, making a total of 72 trials. Of the 24 words presented once only, 4 were presented under each of the six conditions (three types of question × *yes-no*). Similarly, of the 24 words presented twice, 4 were presented under each of the six conditions. When a word was repeated, it always occurred as the 20th item after its first presentation; that is, the lag between first and second presentations was held constant. On its second appearance, the same *type* of question was asked as on the word's first appearance but, for rhyme and sentence questions, a different *specific* question was asked. Thus, when the word TRAIN fell into the rhyme-*yes* category, the question asked on its first presentation might have been "Does the word rhyme with BRAIN?" while on the second presentation the question might have been "Does the word rhyme with CRANE?" For case questions the same question was asked on the two occurrences since each subject was given the same question throughout the experiment (e.g., "Is the word in lowercase?"). This procedure was adopted as early work had shown that subjects' response latencies were greatly slowed if they had to associate *yes* responses to both uppercase and lowercase words.

A constant pool of 48 words was used for all subjects. The words were common concrete nouns. Five presentation formats were constructed in which the words were randomly allocated to the various encoding conditions. Four subjects were tested on each format: Two made *yes* responses with their right hand on the right response key while two used the left-hand key for *yes* responses. The 20 student subjects were paid for their services. They were told that the experiment concerned perception and reaction time; they were warned that some words would occur twice, but they were not informed of the final recall test.

Results and discussion. Response latencies are shown in Table 3. For each subject and each experimental condition (e.g., case–*yes*) the median response latency was calculated for the eight words presented on their first occurrence (i.e., the four words presented only once, and the first occurrence of the four repeated words). The median

TABLE 3
RESPONSE LATENCIES FOR EXPERIMENTS 3 AND 4

Condition	Case	Rhyme	Sentence
1st presentation			
Incidental (Exp. 3)			
Yes	689	816	870
No	705	725	872
Intentional (Exp. 4)			
Yes	687	796	897
No	685	768	911
2nd presentation			
Incidental (Exp. 3)			
Yes	616	689	771
No	634	725	856
Intentional (Exp. 4)			
Yes	609	684	793
No	599	716	866

Note. Mean medians of response latencies are presented.

latency was also calculated for the four repeated words on their second presentation. Only correct responses were included in the calculation of the medians. Table 3 shows the mean medians for the various experimental conditions. There was a systematic increase in response latency from case question to sentence questions. Also, response latencies were more rapid on the word's second presentation—this was especially true for *yes* responses. These observations were confirmed by an analysis of variance. The effect of question type was significant, $F(2, 38) = 14.4$, $p < .01$, but the effect of response type was not ($F < 1.0$). Repeated words were responded to reliably faster, $F(1, 19) = 10.3$, $p < .01$ and the Number of Presentations × Response Type (*yes–no*) interaction was significant, $F(1, 19) = 5.33$, $p < .05$.

Thus, again, deeper level questions took longer to process, but *yes* responses took no longer than *no* responses. The extra facilitation shown by positive responses on the second presentation may be attributable to the greater predictive value of *yes* questions. For example, the second presentation of a rhyme question may remind the subject of the first presentation and thus facilitate the decision.

Figure 3 shows the recall probabilities for words presented once or twice. There is a marked effect of question type (sentence > rhymes > case); retention is again superior for words given an initial *yes* response and recall of twice-presented words is higher than once-presented words. An analysis of variance confirmed these observations. Semantic questions yielded higher recall, $F(2, 38) = 36.9$, $p < .01$; more *yes* responses than *no* responses were recalled, $F(1, 19) = 21.4$, $p < .01$; two presentations increased performance, $F(1, 19) = 33.0$, $p < .01$. In addition, semantically encoded words benefited more from the second presentation, as shown by the significant Question Level × Number of Presentations interaction, $F(2, 38) = 10.8$, $p < .01$.

Experiment 3 thus confirmed that deeper levels of encoding take longer to accomplish and that *yes* and *no* responses take equal encoding times. More important, semantic questions led to higher recall performance and more *yes* response words were recalled than *no* response words. These basic results thus apply as well to recall as they do to recognition. Experiments 1–3 have used an incidental learning paradigm; there are good reasons to believe that the incidental nature of the task is *not* critical for the obtained pattern of results to appear (Hyde & Jenkins, 1973). Nevertheless, it was decided to verify Hyde and Jenkins' conclusion using the present paradigm. Thus, Experiment 4 was a replication of Experiment 3, but with the difference that subjects were informed of the final recall task at the beginning of the session.

Experiment 4

Method. The material and procedures were identical to those in Experiment 3 except that subjects were informed of the final free recall task. They were told that the memory task was of equal importance to the initial phase and that they should thus attempt to remember all words shown in the tachistoscope. A 10-min period was allowed for recall. The subjects were 20 college

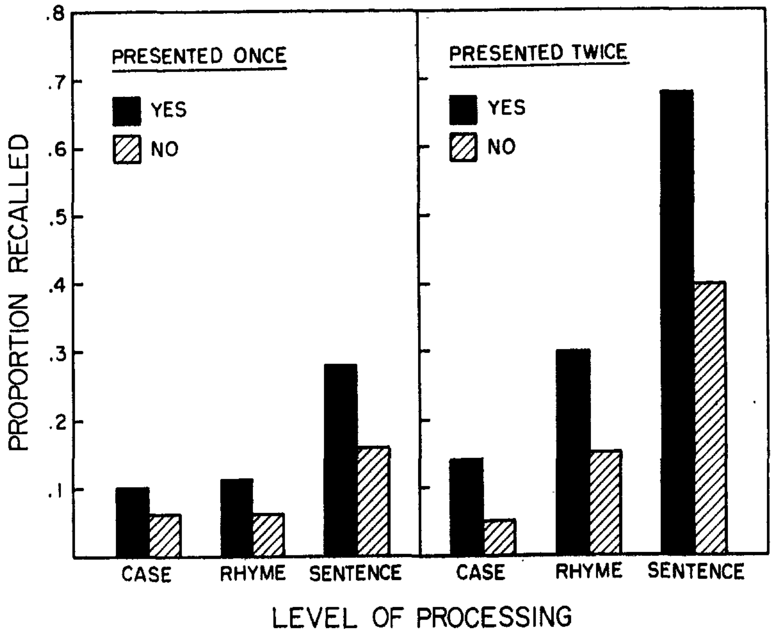

FIGURE 3. Proportion of words recalled as a function of the initial task (Experiment 3).

students, none of whom had participated in Experiments 1, 2, or 3.

Results and discussion. The response latencies are shown in Table 3. These data are very similar to those from Experiment 3, indicating that subjects took no longer to respond under intentional learning instructions. Analysis of variance showed that deeper levels were associated with longer decision latencies, $F(2, 38) = 27.7$, $p < .01$, and that second presentations were responded to faster, $F(1, 19) = 18.9$, $p < .01$. No other effect was statistically reliable.

With regard to the recall results, the analysis of variance yielded significant effects of processing level, $F(2, 38) = 43.4$, $p < .01$, of repetition, $F(1, 19) = 69.7$, $p < .01$, and of response type (*yes–no*), $F(1, 19) = 13.9$, $p < .01$. In addition, the Number of Presentations × Level of Processing interaction, $F(2, 38) = 12.4$, $p < .01$, and the Number of Presentations × Response Type (*yes–no*) interaction, $F(1, 19) = 7.93$, $p < .025$, were statistically reliable. Figure 4 shows that these effects were attributable to superior recall of sentence decisions, twice-presented words and *yes* responses. Words associated with semantic questions and with *yes* responses showed the greatest enhancement of recall after a second presentation.

To further explore the effects of intentional versus incidental conditions more comprehensive analyses of variance were carried out, involving the data from both Experiments 3 and 4. For the latency data, there was no significant effect of the intentional–incidental manipulation, nor did the intentional–incidental factor interact with any other factor. Thus, knowledge of the final recall test had no effect on subjects' decision times. In the case of recall scores, intentional instructions yielded superior performance, $F(1, 38) = 11.73$, $p < .01$, and the Intentional–Incidental × Number of Presentations interaction was significant, $F(1, 38) = 5.75$, $p < .05$. This latter effect shows that the superiority of intentional instructions was greater for twice-presented items. No other interaction involving the incidental-intentional factor was significant. It may thus be concluded that the pattern of results obtained in the present

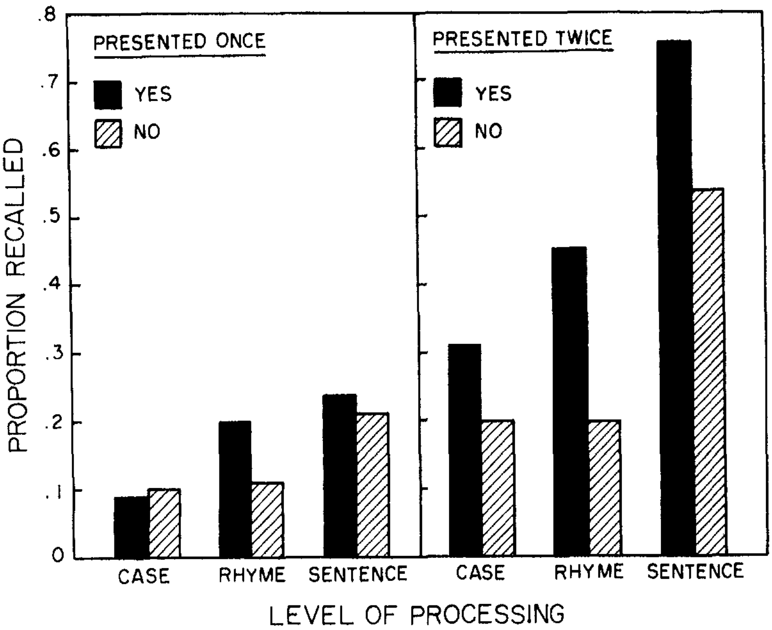

FIGURE 4. Proportion of words recalled as a function of the initial task (Experiment 4).

experiments does not depend critically on incidental instructions.

The findings that intentional recall was superior to incidental recall, but that decision times did not differ between intentional and incidental conditions, is at first sight contrary to the theoretical notions proposed in the introduction to this article. If recall is a function of depth of processing and depth is indexed by decision time, then clearly differences in recall should be associated with differences in initial response latency. However, it is possible that further processing was carried out in the intentional condition, *after* the orienting task question was answered, and was thus not reflected in the decision times.

Discussion of Experiments 1–4

Experiments 1–4 have provided empirical flesh for the theoretical bones of the argument advanced by Craik and Lockhart (1972). When semantic (deeper level) questions were asked about a presented word, its subsequent retention was greatly enhanced. This result held for both recognition and recall; it also held for both incidental and intentional learning (Hyde & Jenkins, 1969, 1973; Till & Jenkins, 1973). The reported effects were both robust, and large in magnitude: Sentence–*yes* words showed recognition and recall levels which were superior to case–*no* words by a factor ranging from 2.4 to 13.6. Plainly, the nature of the encoding operation is an important determinant of both incidental and intentional learning and hence of retention.

At the same time, some aspects of the present results are clearly inconsistent with the depth of processing formulation outlined in the introduction. First, words given a *yes* response in the initial task were better recalled and recognized than words given a *no* response, although reaction times to *yes* and *no* responses were identical. Either reaction time is not an adequate index of depth, or depth is not a good predictor of subsequent retention. We will argue the former case. If depth of processing (defined loosely as increasing semantic–associative analysis of the stimulus) is decoupled from processing time, then on the one hand the independent index of depth has been lost, but on the other hand, the results of Experiments 1–4 can be described in terms of qualitative differences in encoding operations rather than simply in terms of increased processing times. The following section describes evidence relevant to the question of whether retention performance is primarily a function of "study time" or the qualitative nature of mental operations carried out during that time.

The results obtained under intentional learning conditions (Experiment 4) are also not well accommodated by the initial depth of processing notions. If the large differences in retention found in Experiments 1–3 are attributable to different depths of processing in the rather literal sense that only structural analyses are activated by the case judgment task, phonemic analyses are activated by rhyme judgments, and semantic analyses activated by category or sentence judgments, then surely under intentional learning conditions the subject would analyse and perceive the name and meaning of the target word with all three types of question. In this case equal retention should ensue (by the Craik and Lockhart formulation), but Experiment 4 showed that large differences in recall were still found.

A more promising notion is that retention differences should be attributed to *degrees of stimulus elaboration* rather than to differences in depth. This revised formulation retains the important point (borne out by Experiments 1–4) that the qualitative nature of encoding operations is critical for the establishment of a durable trace, but gets away from the notions that semantic analyses necessarily always follow structural analyses and that no meaning is involved in shallow processing tasks.

Discussion of the best descriptive framework for these studies will be resumed after further experiments are reported; for the moment, the term *depth* is retained to signify greater degrees of semantic involvement. Before further discussions of the theoretical framework are presented, the following section describes attempts to evaluate the relative effects of processing time and the qualitative nature of encoding operations on the retention of words.

Processing Time Versus Encoding Operations

As a first step, the data from Experiment 2 were examined for evidence relating the effects of processing time to subsequent memory performance. At first sight, Experiment 2 provided evidence in line with the notion that longer categorization times are associated with higher retention levels—Figure 2 demonstrated linear relationships between initial decision latency and subsequent recognition performance. However, if it is processing time which determines performance, and not the qualitative nature of the task, then *within one task,* longer processing times should be associated with superior memory performance. That is, with the qualitative differences in processing held constant, performance should be determined by the time taken to make the initial decision. On the other hand, if differences in encoding operations are critical for differences in retention, then memory performance should vary between orienting tasks, but within any given task, retention level should not depend on processing time.

This point was explored by analyzing the data from Experiment 2 in terms of fast and slow categorization times. The 10 response latencies for each subject in each condition were divided into the 5 fastest responses and the 5 slowest responses. Next, mean recognition probabilities for the fast and slow subsets of words were calculated across all subjects for each condition. The results of this analysis are shown in Figure 5; mean medians for the response latencies in each subset are plotted against recognition probabilities. If processing *time* were crucial, then the words which fell into the slow subset for each task should have been recognized at higher levels than words which elicited fast responses. Figure 5 shows that this did not happen. Slow responses were recognized little better than fast responses within each level of analysis. On the other hand, the qualitative nature of the task continued to exert a very large effect on recognition performance, suggesting again that it is the nature of the encod-

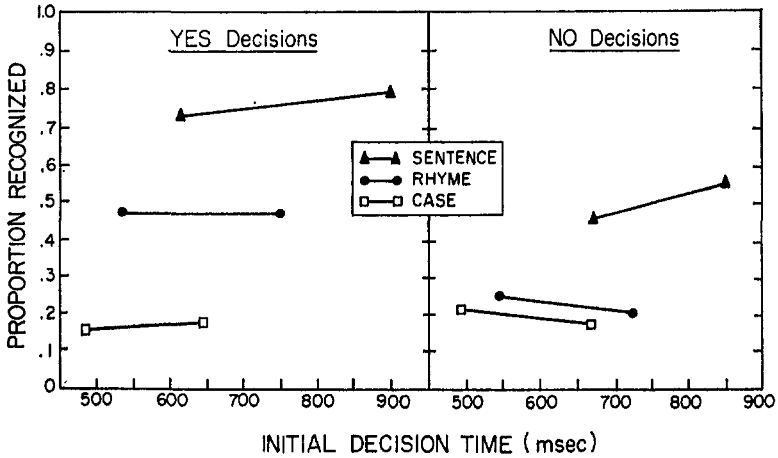

FIGURE 5. Recognition of words as a function of task and initial decision time: Data partitioned into fast and slow decision times (Experiment 2).

ing operations and not processing time which determines memory performance.

For both *yes* and *no* responses, slow case categorization decisions took longer than fast sentence decisions. However, words about which subjects had made sentence decisions showed higher levels of recognition; 73% as opposed to 17% for *yes* responses and 45% as opposed to 17% for *no* responses. No statistical analysis was thought necessary to support the conclusion that task rather than time is the crucial aspect in these experiments. Since the point is an important one, however, a further experiment was conducted to clinch the issue. Subjects were given either a complex structural task or a simple semantic task to perform; it was predicted that the complex structural task would take longer to accomplish but that the semantic task would yield superior memory performance.

Experiment 5

Method. The purpose of Experiment 5 was to devise a shallow nonsemantic task which was difficult to perform and would thus take longer than an easy but deeper semantic task. In this way, further evidence on the relative contributions of processing time and processing depth to memory performance could be obtained. In both tasks, a five-letter word was shown in the tachistoscope for 200 msec and the subject made a *yes–no* decision about the word. The nonsemantic decision concerned the pattern of vowels and consonants which made up the word. Where V = vowel and C = consonant, the word *brain* could be characterized as CCVVC, the word uncle as VCCCV, and so on. Before each nonsemantic trial the subject was shown a card with a particular consonant–vowel pattern typed on it; after studying the card as long as necessary, the subject looked into the tachistoscope and the word was exposed. The experiment was again described as a perceptual, reaction time study concerning different aspects of words and the subject was instructed to respond as rapidly as possible by pressing one of two response keys. The semantic task was the sentence task from previous studies in the series. In this case, the subject was shown a card with a short sentence typed on it; the sentence had one missing word, thus the subject's task was to decide whether the word on the tachistoscope screen would fit the sentence. Examples of sentence-*yes* trials are: "The man threw the ball to the ————" (CHILD) and "Near her bed she kept a ————" (CLOCK). On sentence-*no* trials an inappropriate noun from the general pool was exposed on the tachistoscope. Again the subject responded as rapidly as possible. The subjects were not informed of the subsequent memory test.

The pool of words used consisted of 120 high frequency, concrete five-letter nouns. Each subject received 40 words on the initial decision phase of the task and was then shown all 120 words, 40 targets and 80 distractors mixed randomly, in the second phase. He was then asked to recognize the 40 words he had been shown on the tachistoscope by circling exactly 40 words. Two forms of the recognition test were typed with the same 120 words randomized differently. In all, 24 subjects were tested in the experiment. The pool of 120 words was arbitrarily partitioned into three blocks of 40 words; the first 8 subjects received one block of 40 as targets and the remaining 80 words served as distractors; the second 8 subjects received the second block of 40 words as targets and the third 8 subjects received the third block of 40—in all cases the remaining 80 words formed the distractor pool. Within each group of 8 subjects who received the same 40 target words, 4 received one form of the recognition test and 4 received the other form. Finally, within each group of 4 subjects, each word was rotated so that it appeared (for different subjects) in all four conditions: nonsemantic *yes* and *no* and semantic *yes* and *no*.

Each subject was tested individually. After the two tasks had been explained, he was given a few practice trials, then received 40 further trials, 10 under each experimental condition. The order of presentation of conditions was randomized. After a brief rest period the subject was given the recognition list and told to circle exactly 40 words (those he had just seen on the tachistoscope), guessing if necessary. The subjects were 24 undergraduate students of both sexes, paid for their services.

Results. The results of the experiment are straightforward. Table 4 shows that the nonsemantic task took longer to accomplish but that the deeper sentence task gave rise to higher levels of recognition. Decisions about consonant–vowel structure of words were substantially slower than sentence decisions (1.7 sec as opposed to .85 sec) and this difference was significant statistically, $F(1, 23) = 11.3$, $p < .01$. Neither the response type (yes–no) nor the interaction was significant. For recognition, the analysis of variance showed that sentence decisions gave rise to higher recognition, $F(1, 23) = 40.9$, $p < .001$; *yes* responses were recognized better than *no* responses, $F(1, 23) = 10.6$, $p < .01$, but the Task × Response Type interaction was not significant.

Experiment 5 has thus confirmed the conclusion from the reanalysis of Experiment 2; that it is the qualitative nature of the task —we argue, depth of processing—and not the amount of processing time, which determines memory performance. Figure 2 illustrates that a deep semantic task takes longer to accomplish and yields superior memory performance, but when the two factors are separated it is the task which is crucial, not processing time as such.

One constant feature of Experiments 1–4 has been the superior recall or recognition of words given a *yes* response in the initial perceptual phase. This result has also been reported by Schulman (1974). The reasons for the better retention of *yes* responses are not immediately apparent; for example, it is not obvious that positive responses require deeper processing before the initial perceptual decision can be made. This problem invites a closer investigation of the *yes–no* difference and may perhaps force a further reevaluation of the concept of depth.

TABLE 4

DECISION LATENCY AND RECOGNITION PERFORMANCE FOR WORDS AS A FUNCTION OF THE INITIAL TASK (EXPERIMENT 5)

Response type	Level of processing: Structure	Sentence
	Response latency (sec)	
Yes	1.70	.83
No	1.74	.88
	Proportion recognized	
Yes	.57	.82
No	.50	.69

POSITIVE AND NEGATIVE CATEGORIZATION DECISIONS

Why are words to which positive responses are made in the perceptual-decision task better remembered? As discussed previously, it does not seem intuitively reasonable that words associated with *yes* responses require deeper processing before the decision is made. However, if high levels of retention are associated with "rich" or "elaborate" encodings of the word (rather than deep encodings), the differences in retention between positive and negative words become understandable. In cases where a positive response is made, the encoding question and the target word can form a coherent, integrated unit. This integration would be especially likely with semantic questions: for example, "A four-footed animal?" (BEAR) or "The boy met a ——— on the street" (FRIEND). However, integration of the question and target word would be much less likely in the negative case: "A four-footed animal?"

(CLOUD) or "The boy met a ——— on the street" (SPEECH). Greater degrees of integration (or, alternatively, greater degrees of elaboration of the target word) may support higher retention in the subsequent test. This factor of integration or congruity (Schulman, 1974) between target word and question would also apply to rhyme questions but not to questions about typescript: If the target word is in capital letters (a *yes* decision), the word's encoding would be elaborated no more than if the word had been presented in lowercase type (a *no* decision). This analysis is based on the premise that effective elaboration of an encoding requires further descriptive attributes which (a) are salient, or applicable to the event, and (b) specify the event more uniquely. While positive semantic and rhyme decisions fit this description, negative semantic and rhyme decisions and both types of case decision do not. In line with this analysis is the finding from Experiments 1–4 that while positive decisions are associated with higher retention levels for semantic and rhyme questions, words eliciting positive and negative decisions are equally well retained after typescript judgments.

If the preceding argument is valid, then questions leading to equivalent elaboration for positive and negative decisions should be followed by equivalent levels of retention. Questions which appear to meet the case are those of the type "Is the object bigger than a chair?" In this case both positive target words (HOUSE, TRUCK) and negative target words (MOUSE, PIN) should be encoded with equivalent degrees of elaboration; thus, they should be equally well remembered. This proposition was tested in Experiment 6.

Experiment 6

Method. Eight descriptive dimensions were used in the study: size, length, width, height, weight, temperature, sharpness, and value. For each of these dimensions, a set of eight concrete nouns was generated, such that the dimension was a salient descriptive feature for the words in each set (e.g., size–ELEPHANT, MOUSE; value–DIAMOND, CRUMB). The words were chosen to span the complete range of the relevant dimension (e.g., from very small to very large; very hot to very cold). For each set an additional reference object was chosen such that half of the objects represented by the word set were "greater than" the reference object and half of the objects were "less than" the referent. The reference object was always used in the question pertaining to that dimension; examples were "Taller than a man?" (STEEPLE–*yes*; CHILD–*no*), "More valuable than $10?" (JEWEL–*yes*; BUTTON–*no*). " Sharper than a fork?" (NEEDLE–*yes*; CLUB–*no*). For half of the subjects, the question was reversed in sense, so that words given a *yes* response by one group of subjects were given a *no* response by the other group. Thus, "Taller than a man?" became "Shorter than a man?" (STEEPLE–*no*; CHILD–*yes*).

Each subject was asked questions relating to two dimensions; he thus answered 16 questions—4 yielding positive responses and 4 yielding negative responses for each dimension. Four different versions of the questions and targets were constructed, with two different dimensions being used in each version. Four subjects received each version—two received the original questions (e.g., "heavier than . . ." "hotter than . . .") and two received the questions reversed ("lighter than . . ." "colder than . . ."). Thus each subject received 16 questions; both question type and response type (*yes–no*) were randomized. Subjects were 16 undergraduate students of both sexes; they were paid for their services.

On each trial, the subject looked into a tachistoscope; the question was presented auditorily, and 2 sec later the target word was exposed for 1 sec. The subject responded by pressing the appropriate one of two keys. Subjects were again told that they had to make rapid judgments about words; they were not informed of the retention test. After completing the 16 question trials, subjects were asked to recall the target words. Each subject was reminded of the questions he had been asked. Thus, in this study, memory was assessed in the presence of the original questions.

Results. Again, the results are much easier to describe than the procedure. Words given *yes* responses were recalled with a probability of .36, while words given *no* responses were recalled with a probability of .39. These proportions did not differ significantly when tested by the Wilcoxon test. Thus, when positive and negative decisions are equally well encoded, the respective sets of target words are equally well recalled. The results of this demonstration study suggest that it is not the type of response given to the presented word that is responsible for differences in subsequent recall and recognition, but rather the richness or elaborateness of the encoding. It is possible that negative decisions in Experiments 1–4 were associated with rather poor encodings of the presented words—they did not fit the encoding question and thus did not form an integrated unit with the question. On the other hand, positive responses would be integrated with the question, and thus, arguably, formed more elaborate encodings which supported better retention performance.

Experiment 7 was an attempt to manipulate encoding elaboration more directly. Only semantic information was involved in this study. All encoding questions were sentences with a missing word; on half of the trials the word fitted the sentence (thus all queries were congruous in Schulman's terms). The degree of encoding elaboration was varied by presenting three levels of sentence complexity, ranging from very simple, spare sentence frames (e.g., "He dropped the ———") to complex, elaborate frames (e.g., "The old man hobbled across the room and picked up the valuable ——— from the mahogany table"). The word presented was WATCH in both cases. Although the second sentence is no more predictive of the word, it should yield a more elaborate encoding and thus superior memory performance.

Experiment 7

Method. Three levels of sentence complexity were used: simple, medium, and complex. Each subject received 20 sentence frames at each level of complexity; within each set of 20 there were 10 *yes* responses and 10 *no* responses. The 60 encoding trials were randomized with respect to level of complexity and response type. A constant pool of 60 words was used in the experiment, but two completely different sets of encoding questions were constructed. Words were randomly allocated to sentence level and response type in the two sets (with the obvious constraint that *yes* and *no* words clearly fitted or did not fit the sentence frame, respectively). Within each set of sentence frames, two different random presentation orders were constructed. Five subjects were presented with each format thus generated and 20 subjects were tested in all.

The words used were common nouns. Examples of sentence frames used are: simple, "She cooked the ———" "The ——— is torn"; medium, "The ——— frightened the children" and "The ripe ——— tasted delicious"; complex, "The great bird swooped down and carried off the struggling ———" and "The small lady angrily picked up the red ———." The sentence frames were written on cards and given to the subject. After studying it he looked into the tachistoscope with one hand on each response key. After a ready signal the word was presented for 1.0 sec and the subject responded *yes* or *no* by pressing the appropriate key. The words were exposed for a longer time in this study since the questions were more complex. Subjects were again told that the experiment was concerned with perception and speed of reaction and that they should thus respond as rapidly as possible. No mention was made of a memory test. The 20 subjects were tested individually. They were undergraduate students of both sexes, paid for their services.

After completing the 60 encoding trials, subjects were given a short rest and then asked to recall as many words as they could from the first phase of the experiment. They were given 8 min for free recall. After a further rest, they were given the deck of cards containing the original sentence frames (in a new random order) and asked to recall the word associated with each sentence. Thus there were two retention tests in this study: free recall followed by cued recall.

Results. Figure 6 shows the results. For free recall, there is no effect of sentence complexity in the case of *no* responses, but a systematic increase in recall from simple to complex in the case of *yes* responses. The provision of the sentence frames as cues did not enhance the recall of *no* responses, but had a large positive effect on the recall of *yes* responses; the effect of sentence complexity was also amplified in cued recall. These observations were con-

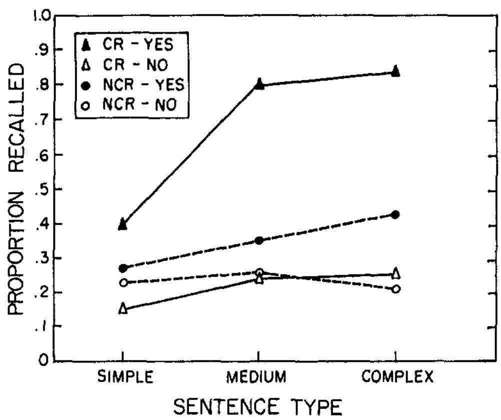

FIGURE 6. Proportion of words recalled as a function of sentence complexity (Experiment 7). (CR = cued recall, NCR = noncued recall.)

firmed by analysis of variance. In free recall, a greater proportion of words given positive responses were recalled than those given negative responses, $F(1, 19) = 18.6$, $p < .001$; the overall effect of complexity was not significant, $F(2, 38) = 2.37$, $p >$ .05, but the interaction between complexity and *yes–no* was reliable, $F(2, 38) = 3.78$, $p < .05$. A further analysis, involving positive responses only, showed that greater sentence complexity was reliably associated with higher recall levels, $F(2, 38) = 4.44$, $p < .025$. In cued recall, there were significant effects of response type, $F(1, 19) = 213$, $p < .001$, complexity, $F(2, 38) = 49.2$, $p < .001$, and the Complexity $\times$ Response Type interaction, $F(2, 38) = 19.2$, $p < .001$. An overall analysis of variance, incorporating both free and cued recall, was also carried out and this analysis revealed significantly higher performance for greater complexity, $F(2, 38) = 36.5$, $p < .001$, for positive target words, $F(1, 19) = 139$, $p < .001$, and for cued recall relative to free recall, $F(1, 19) = 100$, $p <$ .001. All the interactions were significant at the $p < .01$ level or better; the description of these effects is provided by Figure 6.

Experiment 7 has thus demonstrated that more complex, elaborate sentence frames do lead to higher recall, but only in the case of positive target words. Further, the effects of complexity and response type are greatly magnified by reproviding the sentence frames as cues.

These results do not fit the original simple view that memory performance is determined only by the nominal level of processing. In all conditions of Experiment 7 semantic processing of the target word was necessary, yet there were still large differences in performance depending on sentence complexity, the relation between target word and the sentence context, and the presence or absence of cues. It seems that other factors besides the level of processing required to make the perceptual decision are important determinants of memory performance.

The notion of code elaboration provides a more satisfactory basis for describing the results. If a presented word does not fit the sentence frame, the subject cannot form a unified image or percept of the complete sentence, the memory trace will not represent an integrated meaningful pattern, and the word will not be well recalled. In the case of positive responses, such coherent patterns can be formed and their degree of cognitive elaborateness will increase with sentence complexity. While increased elaboration by itself leads to some increase in recall (possibly because richer sentence frames can be more readily recalled) performance is further enhanced when part of the encoded trace is reprovided as a cue. It is well established that cuing aids recall, provided that the cue information has been encoded with the target word at presentation and thus forms part of the same encoded unit (Tulving & Thomson, 1973). The present results are consistent with the finding, but may also be interpreted as showing that a cue is effective to the extent that the cognitive system can encode the cue and the target as a congruous, integrated unit. Elaborate cues by themselves do not aid performance even if they were presented with the target word at input, as shown by the poor recall of negative response words. It is also necessary that the target and the cue form a coherent, integrated pattern.

Schulman (1974) reported results which are essentially identical to the results of Experiment 7. He found better recall of congruous than incongruous phrases; he also found that cuing benefited congruously encoded words much more than incongruous words. Schulman suggests that congruent words can form a relational encoding with their context, and that the context can then serve as an effective redintegrative cue at recall (Begg, 1972; Horowitz & Prytulak, 1969). In these terms, Experiment 7 has added the finding that the semantic richness of the context benefits congruent encodings but has no effect on the encoding of incongruous words.

Is the concept of depth still useful in describing the present experimental results, or are the findings better described in terms of the "spread" of encoding where spread refers to the degrees of encoding elaboration or the number of encoded features? These questions will be taken up in the general discussion, but in outline, we believe that depth still gives a useful account of the major qualitative shifts in a word's encoding (from an analysis of physical features through phonemic features to semantic properties). Within one encoding domain, however, spread or number of encoded features may be better descriptions. Before grappling with these theoretical issues, three final short experiments will be described. The findings from the preceding experiments were so robust that it becomes of interest to ask under what conditions the effects of differential encoding *disappear.* Experiments 8, 9, and 10 were attempts to set boundary limits on the phenomena.

Further Explorations of Depth and Elaboration

The three studies described in this section were undertaken to examine further aspects of depth of processing and to throw more light on the factors underlying good memory performance. The first experiment explored the idea that the critical difference between case-encoded and sentence-encoded words might lie in the similarity of encoding operations within the group of case-encoded words. That is, each case-encoded word is preceded by the same question, "Is the word in capital letters?", whereas each rhyme-encoded and sentence-encoded word has its own unique question. At retrieval, it is likely that the subject uses what he can remember of the encoding question to help him retrieve the target word. Plausibly, encoding questions which were used for many target words would be less effective as retrieval cues since they do not uniquely specify one encoded event in episodic memory. This overloading of retrieval cues would be particularly evident for case-encoded words. It is possible to extend the argument to rhyme-encoded words also; although each target word receives a different rhyme question, phonemic differences may not be so unique or distinctive as semantic differences (Lockhart, Craik, & Jacoby, 1975).

Some empirical support for these ideas may be drawn from two unpublished studies by Moscovitch and Craik (Note 1). The first study used the same paradigm as the present series and compared cued with noncued recall, where the cues were the original encoding questions. It was found that cuing enhanced recall, and that the effect of cuing was greater with deeper levels of encoding. Thus the encoding questions do help retrieval, and their beneficial effect is greatest with semantically encoded words. The second study showed that when several target words shared the same encoding question (e.g., "Rhymes with train?" BRAIN, CRANE, PLANE; "Animal category?" LION, HORSE, GIRAFFE), the sharing manipulation had an adverse effect on cued recall. Further, the adverse effect was greatest for deeper levels of encoding, suggesting that the normal advantage to deeper levels is associated with the uniqueness of the encoded question–target complex, and that when this uniqueness is removed, the mnemonic advantage disappears.

These ideas and findings suggest an experiment in which a case-encoded word is made more unique by being the one word in an encoding series to be encoded in this way. In this situation the one case word might be remembered as well as a word, which, nominally, received deeper processing. Such an experiment in its extreme form would be expensive to conduct, in that one word forms the focus of interest. Experiment 8 pursues the idea of uniqueness in a less extreme form. Three groups of subjects each received 60 encoding trials; each trial consisted of a case, rhyme, or category question. However, each group of subjects received a different number of trials of each question type: either 4 case, 16 rhyme, and 40 category trials; 16, 40, and 4 trials; or 40, 4, and 16 trials, respectively. The prediction was that while the typical pattern of results would be found when 40 trials of one type were given, subsequent recognition performance would be enhanced with smaller set sizes; this enhancement would be especially marked for the case level of encoding.

TABLE 5

Design and Results of Experiment 8

Experimental condition	Case		Rhyme		Category	
	Yes	No	Yes	No	Yes	No
Design: Number of trials per condition						
Group 1	2	2	8	8	20	20
Group 2	8	8	20	20	2	2
Group 3	20	20	2	2	8	8
Proportion recognized						
Group 1	.50	.36	.73	.47	.88	.70
Group 2	.51	.40	.66	.54	.95	.64
Group 3	.49	.43	.90	.70	.91	.68
Set size 4	.50	.36	.90	.70	.95	.64
Set size 16	.51	.40	.73	.47	.91	.68
Set size 40	.49	.43	.66	.54	.88	.70

Experiment 8

Method. Three groups of subjects were tested. Group 1 received 4 case questions, 16 rhyme questions, and 40 category questions. Group 2 received 16, 40, and 4, respectively, while Group 3 received 40, 4, and 16, respectively. At each level of encoding, half of the questions were designed to elicit *yes* responses and half *no* responses. Thus each group received 60 trials; question type and response type were randomized. The design is shown in Table 5.

The subjects were tested individually. Each question was read by the experimenter while the subject looked in the tachistoscope; the word was exposed for 200 msec and the subject responded by pressing one of two response keys. The subjects were informed that the test was a perceptual–reaction time task; the subsequent memory test was not mentioned. After completing the 60 encoding trials, each subject was given a sheet containing the 60 target words plus 120 distractors. He was told to check exactly 60 words—those words he had seen on the tachistoscope.

The same pool of 60 common nouns was used as targets throughout the experiment. Within each experimental group there were four presentation lists; in each case Lists 1 and 2 differed only in the reversal of positive and negative decisions (e.g., category–*yes* in List 1 became category–*no* in List 2). Lsits 3 and 4 contained a fresh randomization of the 60 words, but again Lists 3 and 4 differed between themselves only in the reversal of positive and negative responses. In all, 32 subjects were tested in the experiment; 11 each in Groups 1 and 2, and 10 in Group 3. Two or three subjects were tested under each randomization condition.

Results. Table 5 shows the proportion recognized by each group. Each group shows the typical pattern of results already familiar from Experiments 1–4; there is no evidence of a perturbation due to set size. Table 5 also shows the recognition results organized by set size; it may now be seen that set size does exert some effect, most conspicuously on rhyme–*yes* responses. However, the differences previously attributed to different levels of encoding were certainly not eliminated by the manipulation of set size; in general, when set size was held constant (across groups), strong effects of question type were still found.

To recapitulate, the argument underlying Experiment 8 was that in the standard experiment, the encoding operation for case decisions is, in some sense, always the same; for rhyme decisions, it is somewhat similar from word to word, and is most dissimilar among words in the category task. If the isolation effect in memory (see Cermak, 1972) is a consequence of uniqueness of encoding operations, then when similar encodings (e.g., "case decision" words) are few in number, they should also be encoded uniquely, show the isolation effect, and thus be well recalled. Table 5 shows that reducing the number of case-encoded words from 40 to 4 did not enhance their recall, thus lack of isolation cannot account for their low retention. On the other hand, a reduction in set size did enhance the recall of rhyme-encoded words, thus isolation effects may play some part in these experiments, although they cannot account for all aspects of the results. Finally, it may be of some interest that recall proportions for rhymes–Set Size 4 are quite similar to category–Set Size 40 (.90 and .70 vs. .88 and .70); this observation is at least in line with the notion that when rhyme encodings are made more unique, their recall levels are equivalent to semantic encodings.

Experiment 9: A Classroom Demonstration

Throughout this series of experiments, experimental rigor was strictly observed. Words were exposed for exactly 200 msec; great care was exercised to ensure that subjects would not inform future subjects that a memory test formed part of the experiment; subjects were told that the experiments concerned perception and reaction time; response latencies were painstakingly recorded in all cases. One of the authors, by nature more skeptical than the other, had formed a growing suspicion that this rigor reflected superstitious behavior rather than essential features of the paradigm. This feeling of suspicion was increased by the finding of the typical pattern of results in Experiment 9, which was conducted under intentional learning conditions. Accordingly, a simplified version of Experiment 2 was formulated which violated many of the rules observed in previous studies. Subjects were informed that the main purpose of the experiment was to study an aspect of memory; thus the final recognition test was expected and encoding was intentional rather than incidental. Words were presented serially on a screen at a 6-sec rate; during each 6-sec interval subjects recorded their response to the encoding question. Indeed, the subjects were tested in one group of 12 in a classroom situation during a course on learning and memory; they recorded their own judgments on a question sheet and subsequently attempted to recognize the target words from a second sheet. Reaction times were not measured.

The point of this study was not to attack experimental rigor, but rather to determine to what extent the now familiar pattern of results would emerge under these much looser conditions. If such a pattern does emerge, it will force a further examination of what is meant by deeper levels of processing and what factors underlie the superior retention of deeply processed stimuli.

TABLE 6

Proportion of Words Recognized from Two Replications of Experiment 9

Response type	Case	Rhyme	Category
	1st study		
Yes	.23	.59	.81
No	.28	.33	.62
	2nd study		
Yes	.42	.65	.90
No	.37	.50	.65

Method. On a projection screen, 60 words were presented, one at a time, for 1 sec each with a 5-sec interword interval. All subjects saw the same sequence of words, but different subjects were asked different questions about each word. For example, if the first word was COPPER, one subject would be asked, "Is the word a metal?", a second, "Is the word a kind of fruit?", a third, "Does the word rhyme with STOPPER?", and so on. For each word, six questions were asked (case, rhyme, category × *yes–no*). During the series of 60 words, each subject received 10 trials of each question–response combination, but in a different random order. The questions were presented in booklets, 20 questions per page. Six types of question sheet were made up, each type presented to two subjects. These sheets balanced the words across question types. The subject studied the question, saw the word exposed on the screen, then answered the question by checking *yes* or *no* on the sheet. After the 60 encoding trials, subjects received a further sheet containing 180 words consisting of the original 60 target words plus 120 distractors. The subjects were asked to check exactly 60 words as "old." Two different randomizations of the recognition list were constructed; this control variable was crossed with the six types of question sheets. Thus each of the 12 subjects served in a unique replication of the experiment. Instructions to subjects emphasized that their main task was to remember the words, and that a recognition test would be given after the presentation phase. The materials used are presented in the Appendix.

Result. The top of Table 6 shows that the results of Experiment 9 are quite similar to those of Experiment 2, despite the fact that in the present study subjects knew of the recognition test and words were presented at the rate of 6 sec each. The finding that subjects show exactly the same pattern of results under these very different conditions attests to the fact that the basic phenomenon under study is a robust one. It parallels results from Experiment 4 and previous findings of Hyde and Jenkins (1969, 1973). Before considering the implications of Experiment 9, a replication will be mentioned. This second experiment was a complete replication with 12 other subjects. The results of the second study are also shown in Table 6. Overall recognition performance was higher, especially with case questions, but the pattern is the same.

The results of these two studies are quite surprising. Despite intentional learning conditions and a slow presentation rate, subjects were quite poor at recognizing words which had been given shallow encodings. Since subjects in this experiment were asked to circle exactly 60 words, they could not have used a strict criterion of responding. Thus their low level of recognition performance in the case task must reflect inadequate initial registration of the information or rapid loss of registered information. Indeed, chance performance in this task would be 33%; we have not corrected the data for chance in any experiment. The question now arises as to why subjects do not encode case words to a deeper level during the time after their judgment was recorded. It is possible that recognition of the less well-encoded items is somehow adversely affected by well-encoded items. It is also possible that subjects do not know how best to prepare for a memory test and thus do no further processing of each word beyond the particular judgment that is asked. A third hypothesis, that subjects were poorly motivated and thus simply did not bother to rehearse case words in a more effective way, is put to test in the final experiment. Here subjects were paid by results; in one condition the recognition of case words carried a much higher reward than the recognition of category words.

In any event, Experiment 9 has demonstrated that encoding operations constitute an important determinant of learning or retention under a wide variety of experimental conditions. The finding of a strong effect under quite loosely controlled classroom conditions, without the trappings of timers and tachistoscopes, is difficult to reconcile with the view that was implicit in the initial experiments of the series: that processing of an item is somehow stopped at a particular level and that an additional fraction of a second would have led to better performance. This view is therefore now rejected. It seems to be the qualitative nature of the encoding achieved that is important for memory, regardless of how much time the system requires to reach some hypothetical level or depth of encoding.

Experiment 10

The final experiment to be reported was carried out to determine whether subjects can achieve high recognition performance with case-encoded words if they are given a stronger inducement to concentrate on these items. Subjects were paid for each word correctly recognized; also, they were informed beforehand that a recognition test would be given. Correct recognition of the three types of word was differentially rewarded under three different conditions. Subjects know that case, rhyme, and category words carried either a 1¢, 3¢, or 6¢ reward.

Method. Subjects were tested under the same conditions as subjects in Experiment 9. That is, 60 words were presented for 1 sec each plus 5 sec for the subject to record his judgment. Each subject had 20 words under each encoding condition (case, rhyme, category) with 10 *yes* and 10 *no* responses in each condition. As in Experiment 9, each word appeared in each encoding condition across different subjects. After the initial phase, subjects were given a recognition sheet of 180 words (60 targets plus 120 distractors) and instructed to check exactly 60 words.

There were three experimental groups. All subjects were informed that the experiment was a study of word recognition, that they would be paid according to the number of words they recognized, and therefore that they should attempt to learn each word. The groups differed in the value associated with each class of word: Group 1 subjects knew that they would be paid 1¢, 6¢, and 3¢ for case, rhyme, and category words, respectively; Group 2 subjects were paid 3¢, 1¢, and 6¢, respectively; and Group 3 subjects were paid 6¢, 3¢, and 1¢, respectively. These conditions are summarized in Table 7. Thus, across groups, each class of words was associated with each reward. There were 12 undergraduate subjects in each of three groups.

Results. Table 7 shows that while recognition performance was somewhat higher than the comparable conditions of Experiment 9 (Table 6), the differential reward manipulation had no effect whatever. An analysis of variance confirmed the obvious; there were significant effects due to type of encoding, $F(2, 22) = 90.7$, $p < .01$, response type (*yes–no*), $F(1, 11) = 42.4$, $p < .01$, and the Encoding × Response Type interaction, $F(2, 22) = 4.13$, $p < .05$, but no significant main effect or interactions involving the differential reward conditions.

TABLE 7

Proportions of Words Recognized Under Each Condition in Experiment 10

Encoding operation	Reward value 1 cent	3 cents	6 cents	*M*
Case				
Yes	.50	.51	.54	.52
No	.51	.50	.52	.51
Rhyme				
Yes	.73	.73	.69	.72
No	.53	.50	.60	.54
Category				
Yes	.93	.89	.88	.90
No	.72	.75	.77	.75
Mean				
Yes	.72	.71	.70	.71
No	.59	.58	.63	.60

Although this experiment yielded a null result, its results are not without interest. Even when subjects were presumably quite motivated to learn and recognize case-encoded words, they failed to reach the performance levels associated with rhyme or category words. Subjects in Group 3 (6–3–1) reported that although they really did attempt to concentrate on case words, the category words were somehow "simply easier" to recognize in the second phase of the study.

Thus, Experiments 8, 9, and 10, conducted in an attempt to establish the boundary conditions for the depth of processing effect, failed to remove the strong superiority originally found for semantically encoded words. The effect is not due to isolation, in the simple sense at least (Experiment 8), it does not disappear under intentional learning conditions and a slow presentation rate (Experiment 9), and it remains when subjects are rewarded more for recognizing words with shallower encodings (Experiment 10). The problem now is to develop an adequate theoretical context for these findings and it is to this task that we now turn.

General Discussion

The experimental results will first be briefly summarized. Experiments 1–4 showed that when subjects are asked to make various cognitive judgments about words exposed briefly on a tachistoscope, subsequent memory performance is strongly determined by the nature of that judgment. Questions concerning the word's meaning yielded higher memory performance than questions concerning either the word's sound or the physical characteristics of its printed form. Further, positive decisions in the initial task were associated with higher memory performance (for more semantic questions at least) than were negative decisions. These effects were shown to hold for recognition and recall under incidental and intentional memorizing conditions. One analysis of Experiment 2 showed that recognition increased systematically with initial categorization time, but a further analysis demonstrated that it was the nature of the encoding operations which was crucial for retention, not the amount of time as such. Experiment 5 confirmed that conclusion. Experiments 6 and 7 explored possible reasons for the higher retention of words given positive responses; it was argued that encoding elaboration provided a more satisfactory description of the results than depth of encoding. Experiment 8 showed that isolation effects could not by themselves give an account of the results, Experiment 9 demonstrated that the main findings still occurred under much looser experimental conditions, and Experiment 10 showed that the pattern of results was unaffected when differential rewards were offered for remembering words associated with different orienting tasks.

This set of results confirms and extends the findings of other recent investigations, notably the series of studies by Hyde, Jenkins, and their colleagues (Hyde, 1973; Hyde and Jenkins, 1969, 1973; Till & Jenkins, 1973; Walsh & Jenkins, 1973) and by Schulman (1971, 1974). It is abundantly clear that what determines the level of recall or recognition of a word event is not intention to learn, the amount of effort involved, the difficulty of the orienting task, the amount of time spent making judgments about the items, or even the amount of rehearsal the items receive (Craik & Watkins, 1973); rather it is the qualitative nature of the task, the kind of operations carried out on the items, that determines retention. The problem now is to develop an adequate theoretical formulation which can take us beyond such vague statements as "meaningful things are well remembered."

Depth of Processing

Craik and Lockhart (1972) suggested that memory performance depends on the depth to which the stimulus is analyzed. This formulation implies that the stimulus is processed through a fixed series of analyzers, from structural to semantic; that the system stops processing the stimulus once the analysis relevant to the task has been carried out, and that judgment time might serve as an index of the depth reached and thus of the trace's memorability.

These original notions now seem unsatisfactory in a number of ways. First, the postulated series of analyzers cannot lie on a continuum since structural analyses do not shade into semantic analyses. The modified view of "domains" of encoding (Sutherland, 1972) was suggested by Lockhart, Craik, and Jacoby (1975). The modification postulates that while some structural analysis must precede semantic analysis, a full structural analysis is not usually carried out; only those structural analyses necessary to provide evidence for subsequent domains are performed. Thus, in the case where a stimulus is highly predictable at the semantic level, only rather minimal structural analysis, sufficient to confirm the expectation, would be carried out. The original levels of processing viewpoint is also unsatisfactory in the light of the present empirical findings if it is assumed that *yes* and *no* responses are processed to roughly the same depth before a decision can be made, since there are no differences in reaction times, yet there are large differences in retention of the words.

Second, large differences in retention were also found when the complexity of the encoding context was manipulated. Experiment 7 showed that elaborate sentence frames led to higher recall levels than did simple sentence frames. This observation suggests than an adequate theory must not focus only on the nominal stimulus but must also consider the encoded pattern of "stimulus in context."

Third, and most crucial perhaps, strong encoding effects were found under intentional learning conditions in Experiments 4 and 9; it is totally implausible that, under such conditions, the system stops processing the stimulus at some peripheral level. Unless one assumes complete perversity of subjects, it must be clear that the word is fully perceived on each trial. Thus, differential depth of encoding does not seem a promising description, except in very general terms. Finally, as detailed earlier, initial processing time is not always a good predictor of retention. Many of the ideas suggested in the Craik and Lockhart (1972) article thus stand in need of considerable modification if that processing framework is to remain useful.

Degree of Encoding Elaboration

Is spread of encoding a more satisfactory metaphor than depth? The implication of this second description is that while a verbal stimulus is usually identified as a particular word, this minimal core encoding can be elaborated by a context of further structural, phonemic, and semantic encodings. Again, the memory trace can be conceptualized as a record of the various pattern-recognition and interpretive analyses carried out on the stimulus and its context; the difference between the depth and spread viewpoints lies only in the postulated organization of the cognitive structures responsible for pattern recognition and elaboration, with depth implying that encoding operations are carried out in a fixed sequence and spread leading to the more flexible notion that the basic perceptual core of the event can be elaborated in many different ways. The notion of encoding domains suggested by Lockhart, Craik, and Jacoby (1975) is in essence a spread theory, since encoding elaboration depends more on the breadth of analysis carried out within each domain than on the ordinal position of an analysis in the processing sequence. However, while *spread* and *elaboration* may indeed be better descriptive terms for the results reported in this paper, it should be borne in mind that retention depends critically on the qualitative nature of the encoding operations performed—a minimal semantic analysis is more beneficial for memory than an elaborate structural analysis (Experiment 5).

Whatever the sequence of operations, the present findings are well described by the idea that memory performance depends on the elaborateness of the final encoding. Retention is enhanced when the encoding context is more fully descriptive (Experiment 7), although this beneficial effect is restricted to cases where the target stimulus is compatible with the context and can thus form an integrated encoded unit with it. Thus the increased elaboration provided by complex sentence frames in Experiment 7 did not increase recall performance in the case of negative response words. The same argument can be applied to the generally superior retention of positive response words in all the present experiments; for positive responses the encoding question can be integrated with the target word and a more elaborate unit formed. In certain cases, however, positive responses do not yield a more elaborately encoded unit; such cases occur when negative decisions specify the nature of the attributes in question as precisely as positive decisions. For example, the response *no* to the question "Is the word in capital letters?" indicates clearly that the word is in lowercase letters; similarly a *no* response to the question "Is the object bigger than a man?" indicates that the object is smaller than a man. When *no* responses yield as elaborate an encoding as *yes* responses, memory performance levels are equivalent. There is nothing inherently superior about a *yes* response; retention depends on the degree of elaboration of the encoded trace.

Several authors (e.g., Bower, 1967; Tulving & Watkins, 1975) have suggested that the memory trace can be described in terms of its component attributes. This viewpoint is quite compatible with the notion of encoding elaboration. The position argued in this section is that the trace may be considered the record of encoding operations carried out on the input; the function of these operations is to analyze and specify the attributes of the stimulus. However, it is necessary to add that memory performance cannot be considered simply a function of the number of encoded attributes; the qualitative nature of these attributes is critically important. A second equivalent description is in terms of the "features checked" during encoding. Again, a greater number of features (especially deeper semantic features) implies a more elaborate trace.

Finally, it seems necessary to bring in the principle of integration or congruity for a complete description of encoding. That is, memory performance is enhanced to the extent that the encoding question or context forms an integrated unit with the target word. The higher retention of positive decision words in Schulman's (1974) study and in the present experiments can be described in this way. The question immediately arises as to *why* integration with the encoding context is so helpful. One possibility is that an encoded unit is unitized or integrated on the basis of past experience and, just as the target stimulus fits naturally into a compatible context at encoding, so at retrieval, re-presentation of part of the encoded unit will lead easily to regeneration of the total unit. The suggestion is that at encoding the stimulus is interpreted in terms of the system's structured record of past learning, that is, knowledge of the world or "semantic memory" (Tulving, 1972); at retrieval, the information provided as a cue again utilizes the structure of semantic memory to reconstruct the initial encoding. An integrated or congruous encoding thus yields better memory performance, first, because a more elaborate trace is laid down and, second, because richer encoding implies greater compatibility with the structure, rules, and organization of semantic memory. This structure, in turn, is drawn upon to facilitate retrieval processes.

Broader Implications

Finally, the implications of the present experiments and the related work reported by Hyde and Jenkins (1969, 1973), Schulman (1971, 1974) and Kolers (1973a; Kolers & Ostry, 1974) will be briefly discussed. All these studies conform to the new look in memory research in that the stress is on mental operations; items are remembered not as presented stimuli acting on the organism, but as components of mental activity. Subjects remember not what was "out there" but what they *did* during encoding.

In more traditional memory paradigms, the major theoretical concepts were traces and associations; in both cases their main theoretical property was strength. In turn, the subject's performance in acquisition, retention, transfer, and retrieval was held to be a direct function of the strength of associations and their interrelations. The determinants of strength were also well known: study time, number of repetitions, recency, intentionality of the subject, preexperimental associative strength between items, interference by associations involving identical or similar elements, and so on. In the experiments we have described here, these important determinants of the strength of associations and traces were held constant: nominal identity of items, preexperimental associations among items, intralist similarity, frequency, recency, instructions to "learn" the materials, the amount and duration of interpolated activity. The only thing that was manipulated was the mental activity of the learner; yet, as the results showed, memory performance was dramatically affected by these activities.

This difference between the old paradigm and the new creates many interesting research problems that would not readily have suggested themselves in the former framework. For example, to what extent are the encoding operations performed on an event under the person's volitional strategic control, and to what extent are they determined by factors such as context and set? Why are there such large differences between different encoding operations? In particular, why is it that subjects do not, or can not, encode case words efficiently when they are given explicit instructions to learn the words? How does the ability of one list item to serve as a retrieval cue for another list item (e.g., in an A–B pair) vary as a function of encoding operations performed on the pair as opposed to the individual items? The important concept of association as such, the bond or relation between the two items, A and B, may assume a different form in the new paradigm. The classical ideas of frequency and recency may be eclipsed by notions referring to mental activity.

There are problems, too, associated with the development of a taxonomy of encoding operations. How should such operations be classified? Do encoding operations really fall into types as implied by the distinction between case, rhyme, and category in the present experiments, or is there some underlying continuity between different operations? This last point reflects the debate within theories of perception on whether analysis of structure and analysis of meaning are qualitatively distinct (Sutherland, 1972) or are better thought of as continuous (Kolers, 1973b).

Finally, the major question generated by the present approach is what are the encoding operations underlying "normal" learning and remembering? The experiments reported in this article show that people do not necessarily learn best when they are merely given "learn" instructions. The present viewpoint suggests that when subjects are instructed to learn a list of items, they perform self initiated encoding operaions on the items. Thus, by comparing quantitative and qualitative aspects of performance under learn instructions with performance after various combinations of incidental orienting tasks, the nature of learning processes may be further elucidated. The possibility of analysis and control of learning through its constituent mental operations opens up exciting vistas for theory and application.

REFERENCE NOTE

1. Moscovitch, M., & Craik, F. I. M. *Retrieval cues and levels of processing in recall and recognition.* Unpublished manuscript, 1975. (Available from Morris Moscovitch, Erindale College, Mississauga, Ontario, Canada).

REFERENCES

Begg, I. Recall of meaningful phrases. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 431-439.

Bobrow, S. A., & Bower, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology,* 1969, *80,* 55–61.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press, 1967.

Bower, G. H., & Karlin, M. B. Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology,* 1974, *103,* 751-757.

Broadbent, D. E. *Behaviour.* London: Eyre & Spottiswoode, 1961.

Cermak, L. S. *Human memory: Research and theory.* New York: Ronald, 1972.

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., & Watkins, M. J. The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 599–607.

Eagle, M., & Leiter, E. Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology,* 1964, *68,* 58–63.

Horowitz, L. M., & Prytulak, L. S. Redintegrative memory. *Psychological Review,* 1969, *76,* 519-531.

Hyde, T. S. Differential effects of effort and type of orienting task on recall and organization of highly associated words. *Journal of Experimental Psychology,* 1973, *79,* 111–113.

Hyde, T. S., & Jenkins, J. J. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology,* 1969, *82,* 472–481.

Hyde, T. S., & Jenkins, J. J. Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 471–480.

Jacoby, L. L. Test appropriate strategies in retention of categorized lists. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 675–682.

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

Kolers, P. A. Some modes of representation. In P. Pliner, L. Krames, & T. Alloway (Eds.), *Communication and affect: Language and thought.* New York: Academic Press, 1973. (b)

Kolers, P. A., & Ostry, D. J. Time course of loss of information regarding pattern analyzing operations. *Journal of Verbal Learning and Verbal Behavior,* 1974, *13,* 599–612.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. L. Depth of processing in recognition and recall: Some aspects of a general memory system. In J. Brown (Ed.), *Recognition and recall.* London: Wiley, 1975.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Norman, D. A. (Ed.). *Models of human memory.* New York: Academic Press, 1970.

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

Postman, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964.

Rosenberg, S., & Schiller, W. J. Semantic coding and incidental sentence recall. *Journal of Experimental Psychology,* 1971, *90,* 345–346.

Schulman, A. I. Recognition memory for targets from a scanned word list. *British Journal of Psychology,* 1971, *62,* 335–346.

Schulman, A. I. Memory for words recently classified. *Memory & Cognition,* 1974, *2,* 47–52.

Sheehan, P. W. The role of imagery in incidental learning. *British Journal of Psychology,* 1971, *62,* 235–244.

Sutherland, N. S. Object recognition. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 3). New York: Academic Press, 1972.

Till, R. E., & Jenkins, J. J. The effects of cued orienting tasks on the free recall of words. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 489–498.

Treisman, A., & Tuxworth, J. Immediate and delayed recall of sentences after perceptual processing at different levels. *Journal of Verbal Learning and Verbal Behavior,* 1974, *13,* 38-44.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

Tulving, E., & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review,* 1973, *80,* 352–373.

Tulving, E., & Watkins, M. J. Structure of memory traces. *Psychological Review,* 1975, *82,* 261–275.

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.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review,* 1965, *72,* 89–104.

Wickelgren, W. A. The long and the short of memory. *Psychological Bulletin,* 1973, *80,* 425–438.

(Received February 5, 1975)

Appendix

Each subject in Experiment 9 received the same 60 words in the same order, but six different "formats" were constructed, such that all six possible questions (case, rhyme, category × *yes–no*) were asked for each word (Table A1). Thus, for SPEECH, the questions were (a) Is the word in capital letters? (b) Is the word in small print? (c) Does the word rhyme with each? (d) Does the word rhyme with tense? (e) Is the word a form of communication? (f) Is the word something to wear? Each format contained 10 questions of each type. Negative questions were drawn from the pool of unused questions in that particular format.

TABLE A1

WORDS AND QUESTIONS USED IN EXPERIMENT 9

Word	Rhyme question	Category question	Word	Rhyme question	Category question
SPEECH	each	a form of communication	BEAR	hair	a wild animal
BRUSH	lush	used for cleaning	LAMP	camp	a type of furniture
CHEEK	teak	a part of the body	CHERRY	very	a type of fruit
FENCE	tense	found in the garden	ROCK	stock	a type of mineral
FLAME	claim	something hot	EARL	pearl	a type of nobility
FLOUR	sour	used for cooking	POOL	school	a type of game
HONEY	funny	a type of food	WEEK	peak	a division of time
KNIFE	wife	a type of weapon	BOAT	rote	a mode of travel
SHEEP	leap	a type of farm animal	PAIL	whale	a type of container
COPPER	stopper	a type of metal	TROUT	bout	a type of fish
GLOVE	shove	something to wear	GRAM	tram	a type of measurement
MONK	trunk	a type of clergy	WOOL	pull	a type of material
DAISY	crazy	a type of flower	CLIP	ship	a type of office supply
MINER	liner	a type of occupation	JUICE	noose	a type of beverage
CART	start	a type of vehicle	POND	wand	a body of water
CLOVE	rove	a type of herb	LANE	pain	a type of thoroughfare
ROBBER	clobber	a type of criminal	NURSE	curse	associated with medicine
MAST	past	a part of a ship	LARK	park	a type of bird
FIDDLE	riddle	a musical instrument	STATE	crate	a territorial unit
CHAPEL	grapple	a type of building	SOAP	rope	a type of toiletry
SONNET	bonnet	a written form of art	JADE	raid	a type of precious stone
WITCH	rich	associated with magic	SLEET	feet	a type of weather
ROACH	coach	a type of insect	RICE	dice	a type of grain
BRAKE	shake	a part of a car	TIRE	fire	a round object
TWIG	big	a part of a tree	CHILD	wild	a human being
GRIN	bin	a human expression	DANCE	stance	a type of physical activity
DRILL	fill	a type of implement	FIELD	shield	found in the countryside
MOAN	prone	a human sound	FLOOR	sore	a part of a room
CLAW	raw	a part of an animal	GLASS	pass	a type of utensil
SINGER	ringer	a type of entertainer	TRIBE	scribe	a group of people