

Interaction Between Encoding and Retrieval Operations in Cued Recall

Ronald P. Fisher and Fergus I. M. Craik
University of Toronto, Canada

Three experiments are described in which the qualitative nature of memorial processing was manipulated at both input (encoding) and output (retrieval). In line with previous studies, it was found that for each type of encoding, retention levels were highest when the same type of information was used as a retrieval cue. However, rhyme encoding followed by a rhyme cue gave lower performance levels than did semantic encoding followed by a semantic cue; also, the advantage of semantic encoding over phonemic encoding increased as the retrieval cue was made more compatible with the encoded trace. It was concluded that the notions of encoding specificity and depth of processing are both necessary to give an adequate description of the results.

Craik and Lockhart (1972) suggested that the memory of an event was a function of the "depth" to which the event was initially processed. In this scheme, deeper levels of processing are those concerned with semantic and associative analyses of the stimulus. Some empirical support for these notions was provided by Craik and Tulving (1975). In their experiments, subjects were asked questions about a series of words in an incidental learning paradigm. Before each word was presented, the subject was asked a question, either about the word's physical features (e.g., "Is the word in capital letters?"), its phonemic features (e.g., "Does the word rhyme with chair?"), or its semantic features (e.g., "Is the word a type of animal?"). By Craik and Lockhart's formulation, the physical, phonemic, and semantic questions induce progressively deeper levels of processing; it was thus expected that the semantic questions would be associated with the highest level of per-

formance in a subsequent memory test. This expectation was upheld with both recall and recognition measures and under both incidental and intentional learning conditions.

Although Craik and Tulving's (1975) results lend strong support to the levels-of-processing approach, there is one major deficiency in that approach if it is to lead to a complete account of memory processes; the deficiency is the absence of any specification of retrieval processes. Tulving (1974) has stressed that memory must be viewed as a joint function of stored information (the memory trace) and information provided to the subject at retrieval (the retrieval cue). So far, the levels-of-processing view has yielded some insights into relations between encoding operations and memory performance but has generated little work on possible interactions between the results of various encoding operations and retrieval processes. Although Craik and Tulving's results were interpreted as reflecting encoding variations, it is logically possible that encoding was equivalent under the various experimental conditions and that the differences arose at retrieval. For example, when subjects are given a free-recall or recognition task, they may typically adopt a semantic retrieval mode; that is, they may retrieve stored events in terms of their semantic characteristics. If

This research was supported by Grant A8261 from the National Research Council of Canada to the second author. The authors are grateful for many helpful discussions of these ideas with colleagues; we are especially grateful to Endel Tulving for his very useful comments on an earlier draft.

Requests for reprints should be sent to Ronald P. Fisher, Department of Psychology, Erindale College, University of Toronto, Mississauga, Ontario, Canada L5L 1C6.

this is so, it is plausible that the relatively poor performance following structural and phonemic decisions in Craik and Tulving's experiments was due to a mismatch between encoding and retrieval operations. This line of reasoning suggests that if subjects were induced to adopt structural or phonemic retrieval modes following structural or phonemic encoding, respectively, performance in these conditions would be brought up to the level of the semantic condition.

One purpose of the present series of experiments was to obtain evidence on this alternative account of Craik and Tulving's (1975) results. The method used was to gain some control of both encoding and retrieval processes through manipulation of the input context (by asking different types of questions or by presenting the target word with a rhyming or associated word) and through manipulation of the output context (by presenting different types of retrieval cues). If the alternative account is valid, then matching the nature of input and output contexts should equate memory performance in all cases.

A related purpose of the present experiments was to explore possible relations between the levels-of-processing notions and the encoding-specificity principle (Tulving & Thomson, 1973). This principle states: "Specific encoding operations performed on what is perceived determine what is stored, and what is stored determines what retrieval cues are effective in providing access to what is stored" (Tulving & Thomson, 1973, p. 369). The levels-of-processing manipulations reported by Craik and Tulving (1975) provide powerful examples of ways in which the nature of stored information may be influenced. In these experiments it seems likely that the qualitative nature of the encoded trace is affected by input conditions and that the memory trace of the encoded word is biased to contain more sensory, phonemic, or semantic information, depending on the questions asked during the encoding phase. If this is so, it follows from the encoding-specificity principle that the most effective retrieval cue will differ in these three encoding conditions. It is thus possible (as outlined above) that

the differential retention observed by Craik and Tulving following different encoding conditions does not primarily reflect differential durability of the traces, as suggested by Craik and Lockhart (1972), so much as differential compatibility of trace and cue information. One possibility to be examined in the present experiments is that at retrieval, differently encoded traces are all still present in the memory system and can potentially give rise to equal levels of retention, provided the optimal retrieval cue is provided. The alternative outcome—superior retention levels for deeper processing even when retrieval cues are optimized—would support the original suggestion by Craik and Lockhart that deeper processing supports superior levels of retention.

A further issue relating encoding specificity to levels of processing will be examined in light of the present experiments; the notion is that the specificity of encoded information, and thus the effectiveness of a given retrieval cue, increases at deeper levels of encoding. This suggestion is made on the basis of several other ideas and assumptions. For example, Moscovitch and Craik (1976) made the case for "deeper" semantic encodings being more distinctive; whereas phonemic encodings are formed from a limited vocabulary of speech sounds, the possibilities for combining and recombining "semantic elements" are very much greater. It follows from this idea that a phonemic retrieval cue will tend to elicit many encoded traces, since the cue information pertains to many traces, whereas information in a semantic retrieval cue relates more specifically to a smaller number of traces. That is, there is greater functional cue overload (Baddeley, 1968; Watkins & Watkins, 1975) for shallow traces. It is thus assumed that the effectiveness of a retrieval cue is positively correlated with the extent to which information in the cue duplicates information in the trace and is negatively correlated with the extent to which information in the target trace duplicates information in other traces. From these background ideas, it is reasonable to suggest that the specificity of encoded information, and thus the effectiveness of a

compatible retrieval cue, should increase at deeper levels of encoding.

A final major theoretical implication of the present results concerns the conceptual difference between the notion of "depth" in the levels-of-processing framework and the "strength" of memory traces as described by Wickelgren (1970) and others. Whereas differences in strength presumably refer to quantitative differences—a stronger memory trace has more of the same mnemonic attributes— Craik and Lockhart (1972) stressed qualitative differences between traces encoded at different levels. If the latter view is correct, interactions should occur between type of encoding and type of retrieval cue; that is, no absolute value of memorability can be given to any particular encoded event, since memory performance will depend on the interaction of trace information and retrieval cue information (Tulving, 1974). In the light of depth-of-processing and encoding-specificity views, it is perfectly plausible to conceive of two encoded traces, one of which yields the higher recall probability in one retrieval environment but the lower recall probability in a second retrieval environment. It is difficult to see how a theory that postulates a fixed value of strength could cope with this result.

These issues were examined by manipulating the qualitative nature of both the trace and the retrieval cue. Experiment 1 examined interactions between type of encoding operation and type of retrieval cue in the incidental learning paradigm reported by Craik and Tulving (1975). Experiments 2 and 3 extended the generality of the findings by using an intentional learning paradigm and by systematically varying the similarity between the encoding context and the retrieval cue.

Experiment 1

In Experiment 1, three types of encoding were induced by asking subjects a question about each word before it was presented. The paradigm is described more fully by Craik and Tulving (1975). In the present case, the three types of questions used were

(a) rhyme (Does the word rhyme with the following word:—?); (b) category (Does the word belong to the following category:—?); and (c) sentence (Does the word fit into the following sentence:—?). In terms of levels of processing, the rhyme encoding represents the shallowest level; there are no theoretical grounds for distinguishing which of (b) and (c) represents the deeper level. In all cases, half of the questions led to a positive response and half to a negative response (that is, on half of the trials the word did not rhyme with the given word, fit the category, or complete the sentence). The combinations of three question types and two response types yielded six different encoding conditions.

Subjects were told that they were participating in a perception/reaction time experiment. At the beginning of each trial they were shown a card with a question printed on it. After studying the question, the subject looked into a tachistoscope. The word was then exposed for .2 sec. On reading the word, the subject immediately pressed one of two telegraph keys to indicate his or her response. After completing 72 such trials, the subject was unexpectedly given a stack of 72 retrieval cue cards and was asked to recall as many of the tachistoscopically exposed words as possible. Each retrieval cue correctly referred to only one item and either repeated the same information that had been presented as a question at input or gave information of a different type. Thus, a word might be encoded in terms of a category but cued by a rhyme.

Method

The subjects were 20 undergraduates of both sexes, who were paid for their services. Each subject was tested individually. The words used here were 72 common concrete nouns, between four and six letters in length. For each word, three retrieval cues were formed—a rhyme, a category, and a sentence frame. The three cue types were orthogonally combined with the six encoding conditions (three Question Types $\times$ Yes/no) to form 18 encoding-retrieval combinations, with each combination occurring four times per subject. Four presentation formats were devised; 5 subjects were tested on each format. In each format, 12 words fell into each of the six en-

Table 1
Proportions of Words Recalled as a Function of Encoding Question and Retrieval Cue (Experiment 1)

Retrieval cue	Encoding question							
	Positive responses				Negative responses			
	Rhyme	Category	Sentence	<i>M</i>	Rhyme	Category	Sentence	<i>M</i>
Rhyme	.40	.43	.29	.37	.26	.28	.28	.27
Category	.15	.81	.46	.47	.10	.28	.11	.16
Sentence	.10	.50	.78	.46	.08	.18	.08	.11
<i>M</i>	.22	.58	.51		.15	.25	.16	

coding conditions. These groups of 12 were further split down into three groups of 4 words that were later cued either by a rhyme, a category, or a sentence.

In Format 1, the 72 words were randomly assigned to the resulting 18 combinations of encoding and retrieval conditions, 4 words to each condition. In Format 2, each word retained the same "level" of encoding and retrieval, but each positive response was changed to a negative response and vice versa. In Format 3, the 72 words were allocated to new encoding-retrieval conditions. Format 4 replicated Format 3, but with response type changed.

At the time of retrieval, the subject was presented with a stack of 72 retrieval cue cards, each of which was a valid cue for one of the 72 words shown in the encoding phase. Of the 72 cues, 24 were rhymes, 24 were categories, and 24 were sentences with 1 word missing. Thus, like the encoding questions, retrieval cues were rhymes, categories, or sentences. It should be stressed that the retrieval cues (with certain exceptions) were *not* identical to the encoding questions for any one subject. Typically, a word was encoded under one condition (e.g., rhymes with brain—TRAIN) and retrieved with a different, but valid, retrieval cue (e.g., form of transport). Similarly, words whose associated questions yielded negative responses in the encoding phase were given valid (that is, correct) cues at retrieval. The three exceptions to the general rule of cues being different from encoding questions were rhyme-positive followed by a rhyme cue, category-positive followed by a category cue, and sentence-positive followed by a sentence cue. In these three cases, the retrieval cue was identical to the encoding question; thus three conditions involved intralist cues and the remaining conditions involved extralist cues.

Results and Discussion

The probabilities of cued recall under the different encoding-retrieval combinations are shown in Table 1. The results are shown separately for positive and negative an-

swers to the initial encoding questions. In the case of positive responses, recall was higher for the semantic encoding conditions (category and sentence) than for the rhyme condition; also, performance was somewhat higher when semantic retrieval cues were presented. An analysis of variance on the data for positive responses in Table 1 showed significant effects due to encoding conditions, $F(2, 38) = 48.5$, $p < .01$, and to retrieval conditions, $F(2, 38) = 3.37$, $p < .05$. However, the most striking feature of the data is the presence of a strong interaction between encoding questions and retrieval cues; this interaction was also statistically reliable, $F(4, 76) = 29.61$, $p < .01$. Thus, for each encoding condition, the highest recall probability was obtained when the retrieval cue was identical to that used during study.

These results have implications for the theoretical issues raised in the introduction. First, when the qualitative nature of retrieval was manipulated by varying the type of retrieval cue provided, recall levels were highest when the retrieval cue was identical to the encoding question. Thus, in line with Tulving and Thomson's (1973) findings, retrieval cues were typically most effective when the encoded trace had been biased at input to contain the same type of information as contained in the cue. However, provision of the encoding question as a cue did not equate recall in all cases; cued recall in the sentence-encoding/sentence-cue condition was approximately twice as high as recall in the rhyme-encoding/

rhyme-cue condition. Given that the same pattern of results was obtained in the present experiment (in which retrieval mode was controlled) and in the experiment reported by Craik and Tulving (1975), the argument that the "levels-of-processing" effects reported by Craik and Tulving were due to a mismatch between encoding and retrieval operations is seriously weakened. When a word was encoded in a rhyme context, a rhyme cue was most effective; but even in this optimal retrieval context, performance was much lower than that exhibited by appropriately cued "deeper" encodings. To describe the pattern of results for positive responses in Table 1 adequately, the notions of depth of processing and encoding-retrieval compatibility are both required (cf. Nelson, Wheeler, Borden, & Brooks, 1974, who also suggest the use of both principles to complement each other). Finally, it may be noted that the data for positive responses are quite incompatible with the view that memory traces differ only in "strength"; performance levels under each encoding condition do not reflect some absolute level but depend strongly on the type of retrieval cue used.

The data shown on the right-hand side of Table 1 are the cued-recall probabilities of words that were encoded in the context of questions that yielded negative responses. Since the retrieval cues were valid for the target words in all cases, they were necessarily extralist cues in all conditions in negative-response questions. The pattern of results is quite different from the pattern for positive responses in Table 1. First, there is no easily interpreted effect of type of encoding question; although the differences between encoding conditions were found by analysis of variance to be statistically reliable, $F(2, 38) = 6.16$, $p < .01$, the sentence condition gave very similar results to the rhyme condition. This attenuation of the levels-of-processing effect in the case of negative responses is similar to the results found under cued-recall conditions in previous experiments (Craik & Tulving, 1975, Experiment 8; Moscovitch & Craik, 1976). The analysis of variance yielded a main effect of cue condition,

$F(2, 38) = 7.49$, $p < .01$, which is due largely to the higher recall of words with rhyme cues. Critically important for the present argument, there was no significant interaction between encoding question and retrieval cue, $F(4, 76) = 1.21$, $p > .25$.

The pattern of results in Table 1 for negative responses is interpreted as showing that when the encoded word and its encoding question are not "congruent" (Craik & Tulving, 1975; Schulman, 1974), the word's encoded trace is largely unaffected by the nature of the context. In this case, cued-recall performance appears to be determined to a large extent by the a priori effectiveness of particular cues to elicit the target words—in the present experiment, rhyme cues were apparently more effective cues than were other cue types. That is, since the encoding context had little effect in biasing the words' encoding, no interaction with cue type occurred. It should be noted, however, that when the positive responses in Table 1 are contrasted with the negative responses, the different pattern of results is not due simply to the presence of intralist cues in the negative diagonal cells for the positive responses. The table shows that each cell for the positive responses has a higher value than the corresponding cell for negative responses. Thus, the congruent contexts in the positive-response data apparently added some encoded information that was later beneficial to cued recall. The increment was relatively slight in the case of rhyme encodings but substantial in most cases with category or sentence encoding. The increment was highest, of course, when the word was cued by its original encoding question (the negative diagonal of the positive responses in Table 1).

Overall, the results of Experiment 1 are similar to findings reported by Craik and Tulving (1975) and Moscovitch and Craik (1976); namely, the provision of more effective retrieval cues was especially beneficial to congruent and to highly elaborated traces. Craik and Tulving (1975, Experiment 7) manipulated the encoding context for single words by providing sentence frames of different degrees of complexity. When free recall of the target words was

compared to cued recall, it was found that reproviding the noncongruent sentence frames as cues did not increase retention above the free-recall level. When the word and the sentence frame were congruent, however, reproviding the frame did boost recall, and the benefit was greatest with the most complex sentences. Moscovitch and Craik reported the related result that the benefit of cued over free recall was greater at deeper levels of processing; again, cues had relatively little effect on noncongruent encodings. Both these sets of data and the results from the present experiment suggest that the effectiveness of a cue depends not only on its presence during encoding but also on its integration with the target event; given that congruence or integration, the cue is most effective when the original encoding was "deep" or elaborate.

In summary, Experiment 1 has shown that the "levels-of-processing" effect demonstrated by Craik and Tulving (1975) is not an artifact caused by a mismatch between encoded information and the retrieval cue; the depth-of-processing effect remains strong (for positive responses at least) when retrieval is biased to be compatible with encoding. The presence of a strong interaction between encoding and cuing conditions provides severe difficulties for strength theories of memory—retention level is not absolute for any encoding but depends critically on the retrieval environment; the present results thus provide further evidence for the arguments against strength theory advanced by Anderson and Bower (1972) and by Tulving and Thomson (1973). The data from negative responses yielded a rather different pattern. In this case, the only major effect was that of cue type. When the word and its context are not congruent, the effect of encoding question and the encoding-retrieval interaction are largely absent. Discussion of further theoretical issues is deferred until the remaining experiments are described.

Experiment 1 embodied two major procedural weaknesses. First, retrieval cues apparently differed in their a priori effectiveness—Table 1 negative-response data show that the rhyme cues used were as-

sociated with generally higher levels of recall than were the semantic cues. Second, the results were somewhat difficult to interpret, owing to the use of both intralist and extralist cues. In Experiment 2, cues were first equated for "retrieval potency" to remove one source of confusion. In Experiment 3, cues were again equated, and in addition, the use of extralist cues was extended.

Experiment 2

The essential purpose of Experiment 2 was to replicate the interaction between type of encoding question and type of retrieval cue in a slightly different paradigm, using materials equated for the a priori effectiveness of the cues.

Method

In the first phase of the study, a pool of 56 words was formed such that each word could be generated with approximately equal facility from a rhyme cue and a semantically associated cue. This was accomplished by starting with a larger pool of 96 words, each of which had a potential rhyme cue and a potential semantic cue. Two sets of booklets were constructed, one containing the 96 rhyme cues and the other containing the 96 associative cues. The booklets were then distributed to 70 undergraduates in a classroom situation; half received the rhyme words with instructions to generate rhymes to each cue word during successive 10-sec intervals, and half of the subjects received the associative cues with instructions to generate semantic associates during the 10-sec intervals. The point, of course, was to determine the probability of generating the original word from the cue. Subjects were encouraged to respond with as many rhymes (or associates) as they could generate within the allotted time. The original word was considered to have been generated if it appeared in any output position. Typically, subjects generated 2–4 words per cue. Finally, 56 words were selected to be used in the experiment proper. These words were chosen so that (a) for each word the probability of generating it from the rhyme cue was roughly equal to the probability of generating it from the associative cue (the average difference between these probabilities, disregarding sign, was .10), and also (b) for the group of words the probabilities of rhyme and associate generation were as close as possible. For the rhyme and association cues as groups, the mean probabilities (and standard deviations) of generating the target words were .16 ($SD = .12$) and .15 ($SD = .13$), respectively.

Thirty-two paid undergraduates were presented with a list of 64 word pairs. Each word pair con-

tained one word in capital letters, the to-be-remembered word, and printed above it in lowercase letters, the context word. This context word either rhymed with, or was associatively related to, the target word as in bat_{CAT} or dog_{CAT} . The subjects were instructed to learn the target words but to use the context words as an aid to remembering the target words (Tulving & Thomson, 1973). The 64 word pairs contained 4 untested primacy buffer items, 56 items to be tested later, and 4 untested recency buffer items. The pairs were presented visually at a 4-sec rate.

Following the last word pair, the test format was explained: The experimenter presented a series of 56 retrieval cues, each of which either rhymed with, or was associatively related to, only 1 of the target words. The relationship of each cue to its target word was made clear on each trial. Thus, for example, dog_{CAT} during encoding was later cued by "associated with dog" (cf. Arbuckle & Katz, 1976, where the cue-target relationship was unclear to the subject at the time of test). Each retrieval cue was presented for 10 sec, during which time the subject attempted to recall the appropriate target word. A response was scored correct only if the appropriate target word was recalled; thus, if the subject wrote down a target word that was not specified by that particular retrieval cue, it was scored as incorrect (such instances were extremely rare, however).

For each subject, half of the target words were presented in a rhyme context and half in an associative context; half of the retrieval cues were rhymes and half were associates. The resulting 4 combinations of Encoding Context $\times$ Retrieval Cue appeared 14 times each per subject. The assignment of target words to experimental conditions was counterbalanced across subjects. When the retrieval cue level matched the encoding-context level (e.g., rhyme encoding and rhyme retrieval) the retrieval cue was identical to the encoding context. Half of the subjects were given the list of target words in one order and half in the reverse order. The test order always matched the presentation order so that study-test lag was constant for all items.

In summary, the encoding of common words was biased by means of context words that either rhymed with, or were associated with, the target words. In contrast to Experiment 1, learning was intentional in the present experiment; also, all words and encoding contexts were congruent. At test, the target words were cued either by the context word (intra-list cues) or by the alternative cue (extra-list cues). The point of the study was to obtain further information on interactions between level of processing and encoding-retrieval compatibility.

Results and Discussion

The results are shown in Table 2. The table shows that while there is an overall superiority of semantic encoding over rhyme encoding and a superiority of seman-

Table 2
Proportions of Words Recalled as a Function of Encoding Context and Retrieval Cue (Experiment 2)

Retrieval cue	Encoding context		<i>M</i>
	Rhyme	Associate	
Rhyme	.26	.17	.21
Associate	.17	.44	.30
<i>M</i>	.21	.30	

tic cues over rhyme cues, both effects are qualified by an interaction between encoding and retrieval conditions. That is, semantic cues are superior for semantic encodings, whereas rhyme cues are superior for rhyme encodings. The reliability of these effects was assessed by means of analysis of variance. The effect of encoding condition was statistically reliable, $F(1, 31) = 16.13$, $p < .01$, as was the effect of retrieval condition, $F(1, 31) = 24.55$, $p < .01$. The interaction between encoding and retrieval conditions was also statistically significant, $F(1, 31) = 80.9$, $p < .01$.

From these results, it is concluded that the compatibility between encoding conditions and cue conditions is a critically important determinant of recall. The present pattern of results thus confirms the findings from Experiment 1 (Table 1) and replicates previous results showing interactions between encoding and retrieval conditions (Arbuckle & Katz, 1976; Brooks, 1974; Tulving & Osler, 1968; Owings, Note 1). However, the present results also show that when encoding and cue conditions are optimally matched, the semantic encoding/semantic cue combination is associated with higher levels of retrieval than is the combination of rhyme encoding and rhyme cue (Brooks, 1974; Nelson et al., 1974). Thus again, both notions of depth of processing (or *type* of encoding) and encoding-cue compatibility are necessary to describe the results. Experiment 3 explores the further case in which both intralist cues and extralist cues within an encoding domain were utilized.

Experiment 3

In Experiment 3, the encoding of each word was again biased to emphasize its semantic or phonemic features by means of an accompanying context word. At retrieval, three different types of cues were used: first, identical cues—the same word that accompanied the target word at input; second, similar cues—in this case the cue was of the same type (for example, a semantic relationship) as that induced at input; third, different cues—in this case retrieval cues represented the alternative relationship to the target words to that emphasized at input. All cues were equated for their prior likelihood of generating the target word. Again, the purpose of Experiment 3 was to explore the possible interactions between depth of processing at input and the compatibility of encoding and retrieval cue at output.

Method

Twenty-four undergraduates were presented with 62 pairs of words, visually, at a 4-sec rate. The subjects were instructed to learn the word in capital letters in each pair but also to pay attention to the accompanying context word in lowercase letters "since it may help you to remember the word." Half of the context words rhymed with the target word and half of the context words were high associates of the target word. Of the 62 words presented, 4 were untested primacy buffer items and 4 were untested recency items. Immediately following presentation of all 62 pairs, the 54 middle items were tested by presenting one of the three types of cue, at the rate of 10 sec per cue. Each of the 54 cues was meant to refer to only one of the test items, and was either the *identical* word present at input, a cue word bearing a *similar* relation (rhyme or associate) to the

target word, or a word bearing a *different* relation to the target word (e.g., rhyme cue for a target word encoded in the presence of an associate). Thus, if the presented pair was HAIL (pail), the identical cue was "rhymes with pail," the similar retrieval cue was "rhymes with bail," and the different cue was "associated with snow." If the encoding context was HAIL (sleet) the identical cue was "associated with sleet," the similar cue was "associated with snow," and the different retrieval cue was "rhymes with bail."

The words and cues were those used in Experiment 2. The three types of cue were equated for potency—that is, for the a priori likelihood of generating the target word from the cue. Thus, there were two types of encoding context (rhyme and associate) and three levels of similarity between the encoding context and the retrieval cue (identical, similar, and different). The two independent variables—encoding context and similarity between encoding and cue—were factorially combined in a within-subjects design. Assignment of target words to experimental condition was counterbalanced across subjects.

Results

Table 3 shows the probability of cued recall under the various experimental conditions. Clearly, the similarity of the cue to the original encoding had an effect on recall; the more similar the cue, the higher the level of recall. Also, there was an effect of depth of processing in that, at each level of similarity between encoding context and cue, the associative encoding yielded the higher level of recall. Finally, there is apparently an interaction between encoding type and similarity of cue to encoding context, since the superiority of associative over rhyme context is greater at higher degrees of similarity. These impressions were confirmed by an analysis of variance; the effects of encoding context, $F(1, 22) = 18.7, p < .01$; similarity of cue to encoding context, $F(2, 44) = 19.1, p < .01$; and the interaction between similarity and encoding context, $F(2, 44) = 6.71, p < .01$, were all statistically reliable.

The results of Experiment 3 thus confirm that both type of encoding and similarity of cue to encoding context play a major role in determining the level of recall. The results also show that the superiority of semantic over phonemic processing is greatest when the identical cue is used; thus, when retrieval conditions are optimized by re-pre-

Table 3
Proportions of Words Recalled as a Function of Encoding Context and Similarity Between Encoding Context and Retrieval Cue

Encoding/retrieval similarity	Encoding context		<i>M</i>
	Rhyme	Associate	
Identical	.24	.54	.39
Similar	.18	.36	.27
Different	.16	.22	.19
<i>M</i>	.19	.37	

senting the context word as a cue, the semantic-phonemic difference is increased, not decreased. This finding again speaks against the possibility that the advantage to semantic encoding lies with its differential compatibility with spontaneously generated retrieval cues. Viewed from a slightly different angle, the interaction shown in Table 3 could be described as showing that the beneficial effects of similarity between encoding context and cue are greatest with deep, semantic encodings.

The interaction between encoding type and cue type persists under conditions of extralist cuing (*similar* and *different* conditions in Table 3). When *similar* rhymes and associates were used as cues, recall was .18 and .36 for the rhyme and association encoding conditions, respectively. When *different* cues were used, the corresponding figures were .16 and .22. These results again disconfirm the idea that semantic encodings are simply "stronger" than phonemic encodings; recall performance depends on both encoding condition and the cue used.

One final point may be made about the results shown in Table 3. The notion that memory performance depends on the reinstatement of the encoding context at retrieval seems generally accepted (Melton, 1963; Tulving & Thomson, 1973); a strong version of this notion might state that reinstatement of the learning context is the *only* important factor, and that performance is a function of how completely the original context is reinstated at retrieval. However, the data in Table 3 do not support the strong version of the reinstatement notion. Although it is obviously difficult to compare differences in similarity of input context to retrieval cue between rhyme and associative encodings, it seems reasonable to conclude that the *identical* cue in the rhyme condition is more like its encoding context than is the *similar* cue in the associative condition. Yet the rhyme-*identical* condition is associated with a substantially lower level (.24) of cued recall than the associate-*similar* condition (.36). An even stronger comparison may be made between the rhyme-*identical* and associate-*different* conditions; the similarity between cue and encoding context

must presumably be stronger in the former case, yet cued-recall levels are virtually identical. These comparisons lead to the conclusion that both the qualitative nature of the encoding and the degree of overlap between encoding context and cue are important determiners of recall. In general, the pattern of results observed in Experiment 3 is similar to that found by Brooks (1974), with the exception that rhymes were relatively better cues in Brooks's study. We believe that this is due to the different methods used to equate the preexperimental potency of the rhyme and associate cues, with the result that Brooks's rhyme cues were relatively more potent—preexperimentally—than ours.

General Discussion

The questions posed in the introduction have mostly been answered during discussion of the separate experiments; however, the points will be restated briefly and their theoretical implications considered. First, the present results again emphasize the need to take qualitative differences into account when describing the phenomena of memory. In the present experiments, as in previous studies (Craig & Tulving, 1975; Hyde & Jenkins, 1969, 1973), qualitative differences in encoding and retrieval operations had large effects on performance. Further, it was found that the retention levels associated with a particular type of encoding were not fixed but depended heavily on the type of retrieval cue used. This observed interaction between trace information and cue information fits well with Tulving's (1974) views of cue-dependent forgetting but is apparently incompatible with a unidimensional strength view of memory. Second, the present results disconfirmed the possibility that the lower levels of retention associated with phonemic and structural encoding observed in previous experiments were due to a mismatch between encoded information and the retrieval operations typically utilized. In the present studies, when retrieval cues were made compatible with the encoded information, differences between phonemic and semantic encodings

were enhanced, not diminished. Whereas the optimal situation for high levels of memory performance appears to involve semantic encodings in conjunction with identical semantic retrieval cues, it was also found that the retention of phonemically encoded items was highest when identical phonemic cues were provided at retrieval. Both the qualitative nature of the encoding and the degree of compatibility between encoding and cue are apparently necessary to give an adequate account of memory processes. No one factor in isolation—the type of encoding, the type of cue, or the compatibility between encoding and cue—is by itself sufficient to describe performance.

Experiment 3 showed that the beneficial effects of similarity between input context and retrieval cue were greater with semantic encodings than with phonemic encodings. This interaction between depth of processing and differences in retrieval conditions was also reported by Moscovitch and Craik (1976). They found that the difference in free-recall levels between phonemic and semantic encodings was increased under cued recall conditions; also, the semantic superiority was increased when each encoded word had its own unique cue compared to the situation in which several words shared each cue. Cued recall and unique links between target word and cue may both be regarded as examples of "good" retrieval conditions, and both were differentially beneficial to semantic encodings. Moscovitch and Craik suggested that deeper, semantic encodings establish a higher potential ceiling on performance and that retrieval conditions determine the extent to which that potential is realized. By this argument, semantic encodings stand to gain more, so to speak, as retrieval conditions improve. In the present experiments, retrieval conditions were manipulated by varying the similarity between encoding and retrieval contexts; it is suggested that these results may also be viewed as examples of differentially effective retrieval conditions tapping the greater potential of semantic encodings to different extents. Finally, in this section, it should be noted that the foregoing discussion applies only to

cases in which the target item is congruent with the context; in cases where the target cannot be readily integrated with the context, manipulations of retrieval conditions have relatively little effect (negative responses in Table 1 of the present study; Craik & Tulving, 1975, Experiment 7; Moscovitch & Craik, 1976, Experiment 1).

What is the reason for the higher levels of retention associated with semantic encodings? Craik and Lockhart (1972) suggested that the traces of deeper encodings decayed less rapidly, that such encodings were more durable in the system. An alternative description, which now seems preferable, is that semantic encodings are typically more distinctive, more discriminable from other memory traces. The importance of uniqueness or "isolation" in episodic memory was stressed by Koffka (1935) in a discussion of Von Restorff's experiments; more recently, the notion that deeper, semantic encodings are more distinctive was advanced by Klein and Saltz (1976). Both Lesgold and Goldman (1973) and Moscovitch and Craik (1976) explicitly manipulated uniqueness and found greater uniqueness in encoding to be associated with higher levels of recall. Further, Moscovitch and Craik found that the beneficial effects of uniqueness were restricted to semantic encodings; uniqueness did not benefit phonemic encodings. This last finding echoes the results and conclusions of Jacoby (1974) and of Nelson and his colleagues (Nelson & Borden, in press; Nelson et al., 1974); these workers suggested that phonemic encodings are less modified by the specific encoding context than are semantic encodings and that phonemic encodings are less distinctive. The greater distinctiveness of semantic encodings will only benefit memory, however, if the modifying context is present at both encoding and at retrieval (Nelson & Borden, in press; Tulving & Thomson, 1973). Two further implications of this view are, first, that if semantic encodings are made more similar to one another, they will be remembered little better than phonemic encodings (Moscovitch & Craik, 1976, Experiment 2); second, that the retrieval of phonemic encoding is less affected than is the retrieval of semantic

encodings by changes in retrieval context (Experiment 3 in the present series).

In conclusion, a final link between notions of levels of processing and encoding specificity is speculatively suggested by the results of Experiment 3. It seems possible, as outlined above, that deeper encodings are more discriminable from other traces; they are relatively unique and contain more specific information and less general, shared information than shallow encodings. If this description is accurate, it follows that shallow encodings may be elicited to some extent by a relatively broad range of retrieval cues, whereas deep encodings require specific cue information for successful retrieval. On the other hand, successful retrieval of a deep encoding will entail a fuller and more adequate reinstatement of the original event. That is, there may be a trade-off between shallow and deep encodings such that the information in shallow encodings can rarely be used to recreate the original percept clearly (due to the similarity of shallow encodings to other traces in the system), but this degree of remembering may be accomplished with the help of a broad range of cues. Deep encodings require more specific cues before the percept is elicited to any extent, but remembering is excellent when that specific cue is provided.

Reference Note

1. Owings, R. A. *Levels of processing, encoding specificity, and retrieval of semantically, phonetically, and graphically encoded words*. Manuscript submitted for publication, 1977.

References

- Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, 79, 97-123.
- Arbuckle, T. Y., & Katz, W. A. Structure of memory traces following semantic and nonsemantic orientation tasks in incidental learning. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 362-369.
- Baddeley, A. D. How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology*, 1968, 20, 249-264.
- Brooks, D. H. *The processing of word features in cued recall*. Unpublished doctoral dissertation, University of South Florida, 1974.
- Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671-684.
- Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, 104, 268-294.
- 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. The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 483-496.
- Klein, K., & Saltz, E. Specifying the mechanisms in a levels-of-processing approach to memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 671-679.
- Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
- Lesgold, A. M., & Goldman, S. R. Encoding uniqueness and the imagery mnemonic in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 193-202.
- Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, 2, 1-21.
- Moscovitch, M., & Craik, F. I. M. Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 447-458.
- Nelson, D. L., & Borden, R. C. Encoding and retrieval effects of dual sensory-semantic cues. *Memory & Cognition*, in press.
- Nelson, D. L., Wheeler, J. W., Jr., Borden, R. C., & Brooks, D. H. Levels of processing and cuing: Sensory versus meaning features. *Journal of Experimental Psychology*, 1974, 103, 971-977.
- Schulman, A. Memory for words recently classified. *Memory & Cognition*, 1974, 2, 47-52.
- Tulving, E. Cue-dependent forgetting. *American Scientist*, 1974, 62, 74-82.
- Tulving, E., & Osler, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 1968, 77, 593-601.
- Tulving, E., & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, 80, 352-373.
- Watkins, O. C., & Watkins, M. J. Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 442-452.
- Wickelgren, W. A. Multitrac strength theory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Received February 8, 1977

Revision received May 19, 1977 ■