PERSPECTIVES ON HUMAN MEMORY AND COGNITIVE AGING:

Essays in Honour of Fergus Craik

edited by

Moshe Naveh-Benjamin

Ben-Gurion University of the Negev

Morris Moscovitch

University of Toronto

Henry L. Roediger, III

Washington University in St. Louis

2002

PSYCHOLOGY PRESS
Taylor & Francis Group

3 CHAPTER

Henry L. Roediger, III
David A. Gallo

Levels of Processing: Some Unanswered Questions

In 1960 Tresselt and Mayzner published an interesting experiment that showed large effects of an independent variable on recall. They gave three groups of subjects 100 words such as "career" and "monopoly" and asked each group to perform a task as they examined each word. One group was told to cross out vowels in each word, a second group was told to copy the words over, and a third group was asked to rate the words on a 7-point scale as to how much they belonged to the concept of "economic." (This rating task made sense in the context of the words selected.) Presentation of the words occurred under incidental learning instructions, so the 10-minute free recall test that occurred later came as a surprise to subjects.

Tresselt and Mayzner's (1960) results are shown in Figure 3.1. Recall was much greater following the "economic" rating task relative to the other two tasks. Recall was also somewhat greater after copying words than after crossing out vowels. The experiment was carefully conducted and there is no reason to doubt the results. However, reading the authors' interpretation of the results some 40 years after publication strikes contemporary researchers as somewhat odd. The economic judgment task improved recall, according to the authors, because it increased "the number of differential responses evoked by these words" relative to the other two conditions, an idea owing to Postman (e.g., Postman, Adams, & Phillips, 1955). Differential responses? No one uses these terms today in the study of human memory.

Levels of Processing

Nowadays researchers attempting to explain results like those in Figure 3.1 would resort to quite different language to account for this striking pattern. We can thank Fergus Craik and Robert Lockhart (1972) for introducing the levels-of-processing framework that provides such explanatory power. Their paper (see also Craik, 1973) proposed that

> trace persistence is a function of depth of analysis, with deeper levels of analysis associated with more elaborate, longer lasting, and stronger traces. Since the organism is normally concerned only with the extraction of meaning from the stimuli, it is advantageous to store the

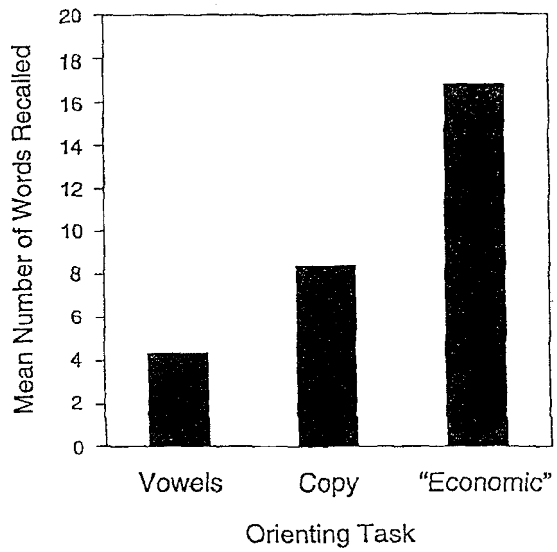

FIGURE 3.1. Mean number of words recalled in each orienting condition of Tresselt and Mayzner (1960).

> products of such deep analyses, but there is usually no need to store the products of preliminary analyses. (p. 675)

In the original levels-of-processing framework, "depth of analysis" referred to basic perceptual processes that ranged from quite superficial (seeing patterns of lines and curves in a visual stimulus such as a word) through intermediate representations (phonemic, based on sound) to the deepest level of semantics or meaning. Depth of processing referred to perception, with memory as a byproduct of perceptual analysis but not (usually) as an end in itself (see Tulving, chapter 2 this volume, for examination of this last idea).

Craik and Lockhart (1972) showed that much of the extant literature could be interpreted by this straightforward but powerful idea of depth of processing. Their paper is one of the most influential papers in the history of cognitive psychology and, indeed, probably one of the most cited papers in all of psychology. Figure 3.2 shows cumulative citations to this paper over the last 30 or so years compared to two other very influential works published the same year (Tulving's [1972] chapter announcing the distinction between episodic and semantic memory and Anderson & Bower's [1972] associative theory of recognition and retrieval processes in free recall). Although all three papers had great impact, like few other papers in the field, Craik and Lockhart's contribution has far outpaced the others in terms of cumulative citations. In fact, we can find no other paper in the past 30 years that has been as influential (and perhaps only Miller's [1956] famous paper rivals it over a longer period, at least within the cognitive psychology of human memory).

Craik and Tulving (1975) developed a paradigm that helped to make the levels-of-processing ideas more concrete and provided the notion that three distinct levels of processing could be distinguished, at least for verbal materials. In considering single words, one can perform a visual analysis of the letters, a phonemic analysis of the word's sound, or a

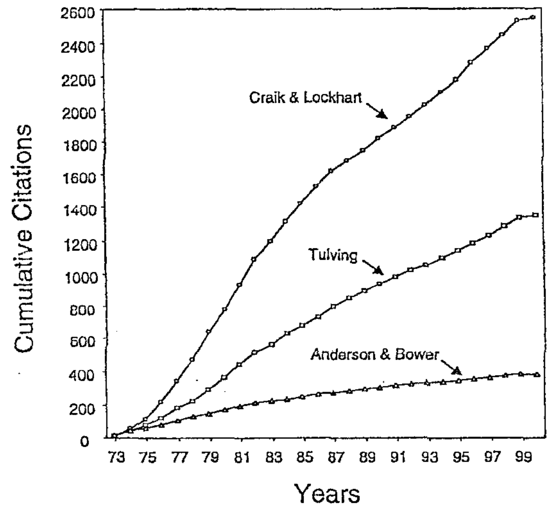

FIGURE 3.2. Cumulative citations from 1973 to May 2000 for Anderson and Bower (1972), Craik and Lockhart (1972), and Tulving (1972).

semantic analysis of its meaning, as shown in Figure 3.3. The paradigm that Craik and Tulving provided, illustrated on the right side of that figure, was designed to effect a particular level of processing by asking subjects questions that would direct attention to a particular level. So, a question about the word's letters would provide a (shallow) visual analysis, one about sound would lead to a phonemic analysis, and a question about a word's category or other meaningful qualities would provoke a semantic (deep) level of processing. The question was assumed to provoke an analysis at the appropriate level and

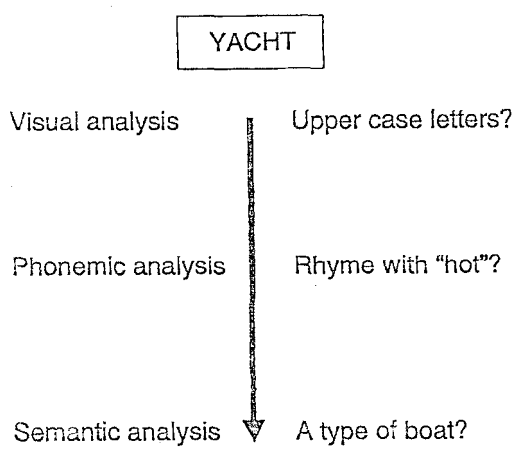

FIGURE 3.3. The standard levels-of-processing framework. During encoding, a word (e.g., "yacht") can be analyzed at several levels (left), with the final level of analysis specified by the type of orienting question (right). Levels of analysis proceed in a more or less linear fashion, with analysis at one level dependent upon completed analyses at shallower levels.

perhaps the levels assumed to occur prior to the critical level. So, for example, a phonemic level of analysis would require analysis of visual features and sound, but not meaning.

In a series of 10 experiments, Craik and Tulving (1975) produced beautiful results showing that free recall and recognition increased directly with the level of processing (as defined by the type of orienting task) accorded words during study. In their first few experiments, they used incidental learning instructions, curtailed the amount of time subjects had to examine words following the question, and meticulously measured time to answer questions (see also Craik, 1973). The assumption seems to have been that these factors were critical in obtaining the levels-of-processing effect. However, in later experiments Craik and Tulving abandoned these methodological strictures and still obtained elegant data. The results shown in Figure 3.4 come from Experiment 9b of their series, in which intentional learning instructions and a 6 s presentation rate were used and the subjects were 12 students in an undergraduate class. Sixty words were studied and students answered yes/no questions about each (case, rhyme, and category), with the answer being "yes" and "no" equally often. The recognition test consisted of the 60 studied words spread amidst 120 lures, with the students instructed to circle exactly 60 words. Despite the relaxed experimental conditions, the data in Figure 3.4 show a dramatic levels of processing effect, which interacted with the yes/no responses during encoding. The "levels" effect was larger for items to which "yes" was the correct answer during the study phase; further, words to which "yes" was the answer were recognized better than those for which "no" was the answer, except for the judgments of case. Of course, in some sense it is no surprise that these effects occurred under slow rates of presentation and intentional learning, because Jenkins and his colleagues had previously shown that type of instruction and time on the task had little effect in similar tasks (Hyde & Jenkins, 1969; Walsh & Jenkins, 1973).

These basic findings were repeated several times in the Craik and Tulving (1975) series of experiments and have been observed in dozens, maybe hundreds, of experiments in

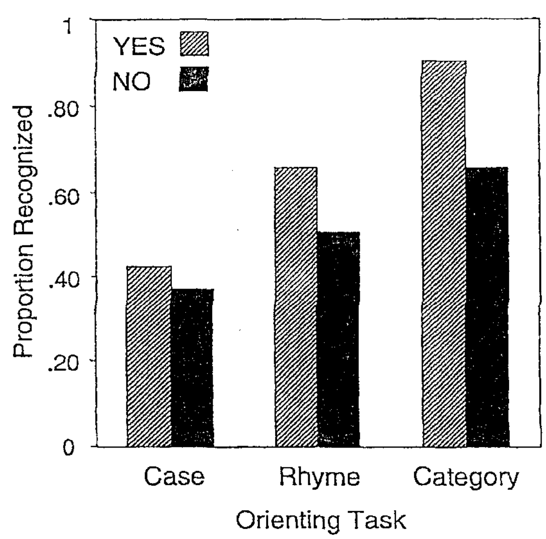

FIGURE 3.4. Mean proportion of words recognized as a function of orienting task and type of question (i.e., "yes" or "no" response) in Craik and Tulving (1975, Experiment 9b).

the past quarter century. Lockhart and Craik (1990) summarized much of the evidence surrounding these phenomena a decade ago and the volume has increased considerably since then. If a researcher today produces data like those of Tresselt and Mayzner (1960) shown in Figure 3.1, we know what to say to explain them: levels of processing is the answer. We would say that the crossing out of vowels involves visual (graphemic) analysis, that the copying of words may involve phonemic/lexical analysis, and that rating the relatedness of a word to the concept of "economic" requires a semantic analysis. No one seems to doubt this answer, and the levels-of-processing framework is mentioned favorably in every textbook (although sometimes qualified as incomplete in certain ways).

But is levels of processing really the answer to why orienting tasks affect performance on memory tests? The purpose of the remainder of this chapter is to raise some unanswered questions about data that arose from research within the levels of processing framework. None of the points should be particularly new, but they may seem new in the current context because the field seems to have collectively forgotten or ignored them.

☐ Unanswered Questions

If the unquestioned answer to the question of "Why do different orienting tasks strongly affect memory performance?" is "levels of processing," then the purpose of the remainder of the chapter is to ask the unanswered question of "Why does anyone believe this statement?" By 1980 we knew it wasn't true. Now, for readers familiar with the experimental psychology literature of the 1970s, the previous sentence may serve as a retrieval cue to call to mind many papers critical of the levels of processing framework in the 1970s. In fact, it was heavily criticized by a large number of people in papers and chapters: Anderson and Reder (1979), Baddeley (1978), Eysenck (1978), Kolers (1979), Morris, Bransford, and Franks (1977), D. L. Nelson (1979), T. O. Nelson (1977), and Tulving (1979), among other people, all voiced strong criticism. Many of these commentators raised cogent points (but see Lockhart & Craik, 1978, 1990, for replies to some of them). One often-voiced criticism had to do with the alleged circularity of the approach, which never worried many of us (see Lockhart, chapter 7 this volume, for discussion), or proposed alternative accounts of the data. However, the basic "levels-of-processing data" themselves were rarely called into question.

We do not want to renew the basic criticisms in the papers above. We instead want to resurrect another set of issues raised by data that were a direct product of the levels of processing approach and therefore seem to be inherently problematic. The data we have in mind are contained mostly in other papers and chapters: Craik and Tulving (1975), Lockhart, Craik, and Jacoby (1976), Moscovitch and Craik (1976), Craik (1977), Fisher and Craik (1977), Chow, Currie, and Craik (1978), and Jacoby, Craik, and Begg (1979). The astute reader will note that the common name here is the honoree of this volume: Fergus I. M. Craik. The mark of a great scientist is to test one's own theory as vigorously as one's critics do. We believe Craik did this in the 1970s.

In the remainder of this chapter, we consider some central aspects of the problematic data produced by Craik and his colleagues that provided challenges to the levels of processing approach. These challenges have never been entirely answered (although Lockhart & Craik, 1978, 1990 addressed some points). To presage, the conclusion we reach is that the original levels of processing approach, taken literally, is undercut by these data from the 1970s and 1980s. We consider six fundamental problems for the levels-of-processing approach, mostly ones introduced by Craik himself.

Question 1. Why does the nature of answers to questions so greatly affect recall and recognition, when they do not affect levels of processing in the task?

This problem appears in full force in Figure 3.4. Answering either "yes" or "no" about the word "yacht" to the question, "Is it a type of boat?" requires subjects to access the meaning of the word. Yet recognition is much greater following "yes" answers than following "no" answers. In general, the levels-of-processing effect is greater for questions to which the answer is "yes" in the typical Craik and Tulving (1975) paradigm. There is no ready explanation within Craik and Lockhart's (1972) theory, but the idea added later was that "richness" or "elaborateness" of encoding was the factor that mattered. Craik and Tulving (Experiments 6 and 7) provided tests of this new idea that were ultimately supporting. They argued that questions answered "yes" generally provided for a more congruous encoding than questions answered "no," and this congruity led to richer, more elaborate encodings that could be used more effectively on the test.

Moscovitch and Craik (1976) manipulated orienting tasks at study and the congruity between encoding and retrieval by providing the questions used at study as retrieval cues during the test. They showed that whether the standard levels of processing effect occurs depends on the nature of retrieval conditions (see also Fisher & Craik, 1977), and they concluded that "other factors may be as important as levels of processing in determining memory performance" (p. 455). In particular, the answer to an orienting question ("yes" or "no"), whether the test employed free or cued recall, and whether cues were unique to a word or shared among words all greatly affected performance. However, the question that these results (and many others) raise is, If all these other factors are important in determining recall and recognition, is the idea of levels of processing needed at all? Can the data be subsumed under another theory? Moscovitch and Craik argued that levels of processing ideas are still needed. We consider this issue more fully in Question 6, after discussing some other issues.

Question 2. Why do levels-of-processing effects occur under intentional learning conditions?

The early experiments by Craik (1973) and Craik and Tulving (1975) generally employed incidental learning conditions and brief tachistoscopic presentation of the words that were to be processed. The assumption was that intentional learning instructions promote meaningful processing, and therefore attempts to induce only shallow levels of processing might be undone by intentional instructions. Keep in mind that depth refers to the level of basic perceptual processing of a word, which can presumably be accomplished in a fraction of a second even by a slow reader, so fast presentation and incidental instructions were thought to be critical to the effect. Yet the general finding is that type of learning instructions does not have much effect on the standard levels-of-processing effect, and this conclusion has been generally accepted.

Consider the experiment of Hyde and Jenkins (1969), who had seven independent groups of subjects engage in free recall of 24-word lists. Their design crossed three types of orienting tasks (checking words for e's, quickly estimating the number of letters, or rating words for their pleasantness on a 7-point scale) with intentional and incidental learning instructions. A seventh group was given standard intentional learning instructions with no orienting task. Within the levels-of-processing framework, pleasantness rating is a deep task, and the other two (involving surface features of the words) are shallow tasks.

TABLE 3.1. Mean percent recalled as a function of learning instruction and orienting task in Hyde and Jenkins (1969).

Orienting task	*Mean recall*
Incidental learning	
Check e's	39%
Estimate letters	41%
Pleasantness rating	68%
Intentional learning	
Check e's	43%
Estimate letters	53%
Pleasantness rating	69%
No task	67%

Hyde and Jenkins's (1969) results are shown in Table 3.1, and indeed, pleasantness rating produced greater recall than did the other two tasks. Further, the intentional learning condition with no orienting task produced about the same level of recall as did the pleasantness rating conditions (whether intentional or incidental). This outcome seems to indicate that an instruction to learn material is tantamount to a deep level of processing manipulation (so long as one is a bit cavalier in accepting the null hypothesis).

But this outcome raises a puzzle: When subjects are given so-called shallow orienting tasks under intentional learning instructions, why doesn't recall increase dramatically? The improvement was negligible in the case of checking e's and modest in the case of estimating letters (see Table 3.1). Why didn't recall approach that of the other intentional conditions when subjects knew they had to learn and remember the material? Was it that subjects did not have enough time? Time seems an unlikely candidate, as Hyde and Jenkins (1969) permitted subjects 2 s/item and, in a second experiment, 4 s/item.

Craik and Tulving (1975, Experiment 4) replicated a strong effect of orienting task under intentional learning conditions. They pointed out that the outcome was at odds with the original notion of levels of processing because

> surely under intentional learning conditions the subject would analyze 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. (p. 279)

Craik and Tulving endorsed the concept of degrees of stimulus elaboration as more appropriate to explain the effects of orienting tasks on recall and recognition than the original levels of processing ideas.

Chow, Currie, and Craik (1978) created conditions that would seem to have produced benefits from intentional learning, if it were possible to do so, following shallow orienting tasks. They used the standard Craik and Tulving (1975) paradigm with intentional learning instructions, but with very slow presentation rates (6 or 12 s/word). For the critical conditions, subjects received special instructions. With one set of instructions the subjects were given intentional learning conditions and a short lecture on the levels of processing effect, with the advice given that after completing the orienting task, it would help their memory if they generated an adjective for each word. With another set, the special task given was to decide whether a word would fit into a sentence frame that was provided. The authors expected that the levels of processing effect would be eliminated under these instructions, because all subjects would process items to a deep level. However, the effect was eliminated only when subjects were given the explicit sentence frame and asked if the word would fit into it, not when they had to generate adjectives. Postman and Kruesi (1977) also showed that intentional learning instructions can benefit performance with shallow encoding tasks, but the levels-of-processing effect remains strong in all experiments with intentional encoding.

It is difficult to understand why intentional learning instructions, with large amounts of study time, do not eliminate (or at least greatly attenuate) the effect of orienting tasks on recall and recognition if the original framework emphasizing perceptual levels of processing is valid. Yet, unless subjects are specifically given a second orienting task that demands semantic processing, the levels effect remains under intentional conditions with slow presentation rates. Does the shallow orienting task short-circuit or curtail subjects' ability to engage voluntarily in meaningful processing? Does the shallow orienting task create some inhibitory process that semantic processing cannot overcome? Whatever the answer, Craik and Tulving's (1975) point seems correct: These results surely pose a problem for the original version of the levels-of-processing framework, one that has never been solved.

Question 3. Why can't the levels-of-processing effect be eliminated by reversing the usual procedure during encoding?

The standard paradigm for studying the effect of orienting tasks is to give people the orienting task ("Does the word contain an e?") first and then give them the stimulus to be judged (hedgehog). The idea is to arrest processing at the prescribed level or (in this case) at a visual level of analysis. However, suppose this usual procedure is reversed and the word is given to the subject for several seconds before the question is given. If intentional learning instructions are used, then according to the Craik and Lockhart (1972) formulation, the levels-of-processing effect should disappear. Surely subjects would process the word to a deep, semantic level before the question is even presented and two seconds would be ample time to do so.

As far as we can tell, this sort of experiment has rarely been tried, because it makes little sense from the levels of processing perspective. However, Craik (1977) directly compared these two ways to manipulate the orienting task, with the procedure outlined in Figure 3.5, under both *intentional* and *incidental learning conditions* (Moeser, 1983, also employed the "reversed" paradigm, but only with incidental instructions). In the standard paradigm, subjects received the question and then the word; in the reversed paradigm, subjects saw the word for 2 s and then, after a full 5 s delay, received the question. Surely meaning would have been extracted from the word after 7 s of study under intentional learning conditions. Nonetheless, as can be seen in Figure 3.6, Craik (1977) still observed a very strong effect of orienting tasks under these "reversed" conditions, although the effect was not as great as when the question was given before the word. As Craik (1977, p. 689) noted, it is "mystifying" that these conditions (intentional learning for 7 s with the word given before the question) did not have a greater impact on performance in the shallow conditions and eliminate the levels-of-processing effect.

The problem posed in this section is essentially the same as in the preceding point: Even when subjects are given ample time and conditions designed to encourage meaningful processing, the effect of orienting tasks—the "levels-of-processing effect"—still occurs. This outcome leads to the conclusion that the original levels-of-processing framework is wrong and that other principles must be brought to bear to explain these phenomena, as Craik and Tulving (1975) and Craik (1977) concluded many years ago.

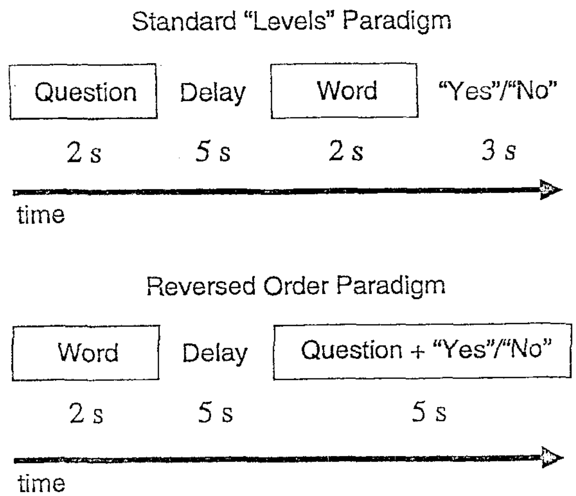

FIGURE 3.5. Representative trials for each of the two methodologies employed by Craik (1977, Experiment II). In the standard paradigm, the orienting question precedes the word, whereas in the reversed paradigm the word precedes the question.

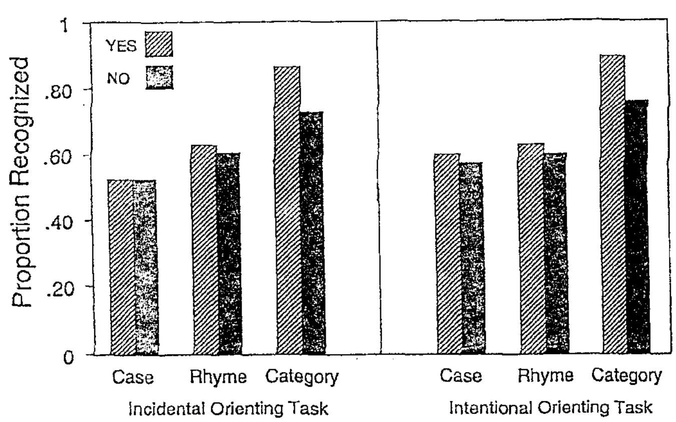

FIGURE 3.6. Mean proportion of words recognized in each learning condition of the reversed order paradigm (i.e., word presented before orienting question), as estimated from Figure 2 of Craik (1977). Performance within each orienting task is presented separately for "yes" and "no" responses to the orienting question.

Question 4. Why do variations within a particular level of processing affect memory performance?

Most experiments in the levels-of-processing tradition manipulate orienting tasks designed to effect distinct levels of processing, much as Craik and Tulving (1975) did. Following Craik and Tulving's lead, many researchers attempted to induce all three levels shown in Figure 3.3 by choosing tasks that require graphemic (visual), phonemic (acoustic), and semantic (meaningful) analyses. Many other researchers contented themselves with two levels of processing, one deep (meaningful) and the other shallow (graphemic or phonemic). The data produced nearly always supported the general levels-of-processing framework, with greater recall or recognition for semantically coded words. However, it is rare that an investigator manipulates the type of orienting task within one domain, visual, phonemics, or semantic. Can recall be similarly affected when orienting tasks are manipulated within one domain (all involving one type or level of processing) as when they are manipulated across domains?

Data from a study by Packman and Battig (1978) indicate that the answer is yes. Packman and Battig gave subjects 50 words under one of seven orienting tasks, manipulated between subjects, and then gave a surprise free recall test after a brief delay. All seven orienting tasks were designed to force a deep level of processing. The full set of data are complex, but consider the results from just three conditions shown in Figure 3.7. The data appear orderly and apparently show a nice levels-of-processing effect. However, all three orienting tasks (rating words for meaningfulness, concreteness, or pleasantness) required meaningful processing. Indeed, the task showing the lowest level of recall of these three was the one that required ratings of meaning! If variations of tasks within one putative level of processing can mimic effects typically obtained between levels of processing, then

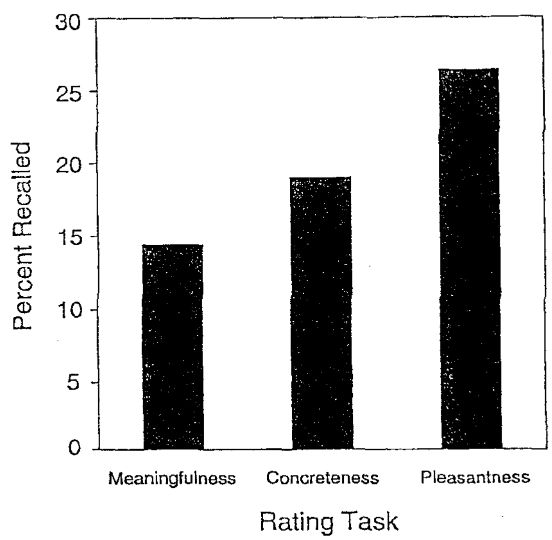

FIGURE 3.7. Mean percent of words recalled in three orienting conditions of Packman and Battig (1978).

we can again question whether the levels-of-processing framework provides a satisfactory explanation of the results from manipulating orienting tasks. Seamon and Virostek (1978) reported rather similar data to those of Packman and Battig, but few other reports exist in which this experimental tactic has been used (see also Challis, Velichkovsky, & Craik, 1996; Hyde & Jenkins, 1973; and Mathews, 1977).

Question 5. If levels-of-processing theory were correct, how could subjects free recall items after orienting tasks that require shallow levels of processing?

Most experiments varying orienting tasks have measured recognition. Recognition memory is usually greater than chance, even after shallow levels of processing (checking for e's or making case judgments). Of course, in recognition memory, researchers have proposed that correct decisions can be made on the basis of familiarity or perceptual fluency (e.g., Mandler, 1980; Jacoby & Dallas, 1981), and these qualities may be supported by traces left from shallow levels of processing. However, the original "levels" framework proposed levels of perceptual analysis. If processing of a word were arrested at graphemic or phonological levels, how could they be recalled minutes or hours later? If all one knew was that a word was in upper case letters or that it had a particular sound, how could the correct word be reconstructed? It could not, of course, yet subjects show reasonable levels of free recall even after shallow orienting tasks.

Roediger, Payne, Gillespie, and Lean (1982) conducted a relevant experiment. Subjects performed a standard set of orienting tasks during the encoding phase of the experiment (graphemic, phonemic, semantic) in a typical Craik–Tulving type of experiment. They studied 60 words under intentional learning conditions, with questions directing attention to one of the three "levels" for 20 words. After list presentation, subjects were distracted by another task for 1 minute to clear short-term memory. Then they were given instructions to recall as many items as possible from the list. A novel feature of the experiment was that two different conditions were employed during testing. All subjects were tested under free recall conditions, but one group was given three 7-minute tests on the words. That is, they recalled as many words as possible for 7 minutes, then their recall sheets were removed and they were given another sheet with instructions to recall as many list items as possible (both ones previously recalled and any new ones that came to mind) for another 7 minutes. Then the procedure was repeated a third time. (The purpose of the experiment was to examine hypermnesia, the recovery of additional information over repeated tests.) A second group of subjects was given one long (21-minute) test, with short breaks after 7 and 14 minutes, during which they were exhorted to continue trying as a control for total time spent in recall. For both groups, subjects drew a line under the last word recalled after each minute, so measures of cumulative recall could be obtained.

The results are shown in Figure 3.8, where solid lines and filled symbols represent the cumulative recall functions for subjects taking three tests and the broken lines and open symbols represent recall of subjects taking one long test. This variable did not affect performance much, replicating prior research by Roediger and Thorpe (1978). Of more interest in the present context is the effect of the orienting task, where the standard levels of processing effect clearly emerged. Further, the recovery of items processed to superficial levels continued to increase over the 21-minute period, which raises the issue noted above: If words were *only* processed to graphemic or phonemic levels, how would they be recalled from long-term memory at all? And how would items that were not semantically coded be recovered over these long intervals? Many researchers assume that free recall is

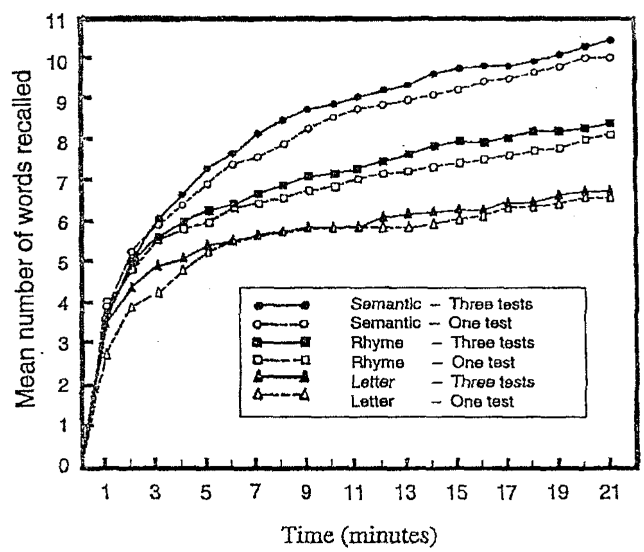

FIGURE 3.8. Cumulative recall curves for sets of items studied under each of the three orienting questions in Roediger et al. (1982). Solid lines and black symbols represent data from the group that received three 7-minute tests, whereas dashed lines and white symbols represent data from the group that received one 21-minute test.

largely driven by semantic or conceptual factors (e.g., Roediger, Weldon, & Challis, 1989); if so, then only words accorded a deep, semantic processing should be recalled.

An alternative idea that would account for the results in Figure 8 and in other free recall experiments is as follows. Perhaps the manipulation of orienting tasks in the standard Craik–Tulving type of encoding experiment does not "arrest" analysis at different levels in the perceptual system. Perhaps all orienting tasks encourage or discourage semantic analysis of words to varying degrees. A task such as pleasantness rating focuses attention on meaning, whereas tasks such as checking for e's or making a case judgment or a rhyme judgment focus attention on dimensions other than meaning. However, for free recall, what matters most is meaningful processing, perhaps because meaning is the most salient dimension upon which words can be differentiated. If we assume that all items recalled during a free recall test were processed for meaning, then the items successfully recalled following graphemic and phonemic encoding were also processed for meaning and not just for graphemic or phonemic features (see Cermak, Schnorr, Buschke, & Atkinson, 1970). Shallow orienting tasks may then be understood in part as divided attention tasks in which subjects' attention is partially diverted from their meaningful analysis of the words.

These statements above echo ideas voiced by Postman and Kruesi (1977) in reflecting on their own experimental results (see also Arbuckle & Katz, 1976). Postman and Kruesi (1977) noted:

> It may be well to remember that subjects come to the experiment with a lifelong habit of processing English words semantically. It may be a basic error to assume that instructions to attend to a nonsemantic property effectively shunt out this disposition. Rather, the imposition of a nonsemantic task may interfere to a greater or lesser degree with a persisting tendency to process words semantically. . . . If one took seriously the assumption that subjects given a

> nonsemantic task no longer engage in semantic processing, one is left with no apparent explanation of why subjects recall as much as they do [following nonsemantic orienting tasks]. (p. 368).

This is just the problem we are raising with reference to the Roediger et al. (1982) data displayed in Figure 3.8, but of course the problem exists in any set of free recall data showing substantial recall (and meaningful clustering of related words) following shallow orienting tasks (e.g., Hyde & Jenkins, 1969). Once again, the challenge proposed by Postman and Kruesi and here has never been successfully answered and offers another reason for doubting the strict (perceptual) levels-of-processing formulation. Other types of tests besides free recall provide another significant challenge, under the rubric of transfer-appropriate processing.

Question 6. Why do orienting tasks affect various tests differently?

The original levels of processing framework represented a relatively pure encoding or trace-dependent (Tulving, 1974) theory of memory. Encoding activities lead to different types of memory traces, these traces support different levels of memory performance and persist differentially over time (Craik & Lockhart, 1972). The implicit assumption was that retrieval processes did not much matter, and neither did the type of test, although Craik and Lockhart (p. 678) did briefly consider such matters. Perhaps the primary empirical problem challenging the original levels of processing formulation in the past 30 or so years has to do with the interaction of encoding and testing conditions. There are several interrelated concerns.

First, in the basic Craik–Tulving variety levels-of-processing experiment, the distinctiveness of the question and the level of processing that the question promotes are confounded. In explaining the fact that positive answers to questions lead to better retention than do negative answers, Craik and Tulving (1975) proposed that the question may serve as a cue during the test, and that "yes" answers produce more congruous encodings than do "no" answers. In addition, the types of questions provide differentially distinctive cues across types of encoding. In asking about the case of words, there are usually only two (upper and lower); in asking about rhymes, the word endings are often somewhat similar across words. However, in asking whether a word belongs to a category or fits into a sentence, the possibilities for dimensions of meaning are many. Moscovitch and Craik (1976) examined the cueing power of questions in three experiments that manipulated type of orienting task and other factors. They concluded that the provision of retrieval cues, the distinctiveness of retrieval cues, and the compatibility of cues with encoded traces all powerfully affected recall, although they thought the concept of levels of processing was necessary to explain their results, too.

Later researchers argued that the concept of levels of processing was superfluous and that notions of transfer-appropriate processing (Morris et al., 1977) or encoding specificity (Tulving, 1979) were the only concepts necessary to explain memory performance in paradigms that manipulate orienting task. The basic type of experiment used as evidence for this proposition was performed by several sets of researchers at roughly the same time: Fisher and Craik (1977); Jacoby (1975); McDaniel, Friedman, and Bourne (1978); and Morris et al. (see also Bransford, Franks, Morris, & Stein, 1979). (The papers focused on somewhat different issues among themselves.) We will present data reported by Morris et al. as illustrative of the main point, but Fisher and Craik (1977) reported rather similar experiments.

Morris et al. (1977, Experiment 1) manipulated the encoding task during study by having subjects read words in sentence frames that were designed to effect either phonemic or semantic encodings. For example, the word "eagle" appeared with a sentence frame such as "_____ rhymes with legal" or "_____ is a large bird" to instantiate the two types of encoding. Subjects responded "yes" or "no" to each item, although for present purposes we will consider only the results from items that were given "yes" responses. Subjects' memories were tested under two different types of retrieval conditions, both employing recognition tests. On a standard yes/no recognition test, subjects were given a series of words that had been studied mixed in with new (nonstudied) words, and their task was to judge each word as studied ("yes") or not ("no"). In this test subjects were assumed to consult the meaning of the words in making their judgments. Other subjects were given a rhyme recognition test in which they also received a long series of test words, none of which had actually been studied. However, half the words on the test rhymed with words that had been studied and subjects were now instructed to say "yes" if the test word (e.g., "beagle") rhymed with a studied word and "no" if it did not.

The experiment represents a 2 × 2 manipulation of encoding and retrieval conditions, as illustrated in Table 3.2. The standard (semantic) recognition test shows the usual levels of processing effect, with recognition following a semantic orienting task (84% hits) much better than performance following the phonemic orienting task (63%). However, on the novel rhyme recognition test, the hit rate was greater following phonemic encoding (49%) than following semantic encoding (33%), a reversal of the typical levels-of-processing effect. Morris et al. (1977) argued that there are no encoding tasks that are inherently deep (or good) for memory. Rather, how well encoded experiences are expressed on tests reflects the appropriateness of the conditions of transfer between study and test (see also Kolers & Roediger, 1984).

This transfer-appropriate processing framework incorporates the levels-of-processing approach as a special case: Semantic processing of words does produce better performance than phonemic or graphemic processing, but only on tests that require access to semantic information. Although most explicit memory tests do rely on conceptual or semantic processing, the rhyme recognition test does not. Fisher and Craik (1977) obtained no levels-of-processing effect on a rhyme-cued recall test and Jacoby (1975) and McDaniel et al. (1978) reported similar results. Certain types of graphemic recall and recognition tests show equal or greater retention following graphemic than semantic processing (e.g., Blaxton, 1989; Stein, 1978).

TABLE 3.2. Transfer-appropriate processing. Encoding conditions emphasized phonemic or semantic dimensions of words. Retrieval conditions required predominant use of one dimension or the other, creating a strong interaction. From Morris, Bransford, and Franks (1977, Experiment 1).

		Retrieval condition	
		Semantic	Rhyme
Encoding condition	Semantic	.84	.33
	Rhyme	.63	.49

Of course, as Lockhart and Craik (1990, p. 101) noted, even when study and test processes are appropriately matched on episodic memory tests, the greatest recall and recognition tends to occur with semantic tasks. For instance, from Table 3.2 it can be seen that, even though the typical "levels" effect was reversed on the rhyme test, performance was still greater with semantic study and semantic test (.84) than with rhyme study and rhyme test (.49). When verbal materials are used, it seems unlikely that any combination of low-level encoding tasks combined with a compatible retrieval task will give as great recall and recognition as do standard semantic encoding and semantic retrieval tests. Therefore, Craik and Lockhart's (1972) emphasis on meaning, especially for verbal materials, as promoting excellent memory performance is well taken (but see Tulving, 1979, for a contrary view).

Although these findings of superiority of semantic encoding and test conditions can be taken as supporting the notion of "levels" of encoding, we feel that this conclusion may be limited. Rather, to us, such an effect serves as a reminder that the to-be-learned materials in almost all levels-of-processing experiments are words. As alluded to earlier, the usual purpose of words in language is to convey semantic information. Given this fact, it is not too surprising that episodic memory for words, qua words, is optimal with an orienting task that does not deflect from the mind's usual way of processing these materials (i.e., semantically) and with a test that is sensitive to meaningful processing (see also Jenkins, 1974, pp. 17–18; Kolers, 1979, pp. 382–384, for related views). In this restricted sense, the levels-of-processing framework is upheld, but again more as a special case of transfer-appropriate processing that applies to memory for words on meaning-based tests. Nonetheless, because meaning is obviously a critical (maybe even the critical) feature in mental representation in most circumstances with verbal materials, the levels-of-processing framework correctly emphasizes this key aspect of mental life, in our opinion.

A case can be made, however, that the underlying dimension promoting good retention is a matter of expertise, not levels of processing (see Bransford et al., 1979). Human adults are expert in using language and in extracting meaning from spoken and written words. If researchers turn their attention to the myriad other types of human activities—baseball hitters' expertise in remembering the kinds of pitches and the pitching patterns of hundreds of pitchers, or ornithologists' memory for the coloration patterns of birds, and on and on endlessly—then the relevant dimension that permits excellent encoding and retrieval might be conceived more broadly as the expertise of the individual for the stimulus vocabulary of their subject matter. Virtually all adult humans are expert in language, hence its use as the lingua franca of memory experiments, but if we broaden consideration to other domains, then meaning may not be the critical feature (Bransford et al., 1979; Kolers & Roediger, 1984).

The problem of generalizability across domains, even within the sphere of language, is also highlighted by the fact that the levels-of-processing effect disappears on perceptual implicit memory tests, as was first shown by Jacoby and Dallas (1981). They conducted a Craik–Tulving type of encoding experiment and measured priming on a word identification test in which subjects were given brief glimpses of words (30 ms or so of exposure time) and tried to guess the identity of the flashed words. Jacoby and Dallas (1981) obtained priming from prior encoding (i.e., subjects identified previously studied words at higher levels than words that had not been studied), but the type of orienting task (graphemic, phonemic, or semantic) did not affect priming on the word identification test. Similar results have been obtained with other perceptual implicit memory tests. Levels of processing generally produces little or no effect on priming on tests such as word stem or word fragment completion in which subjects identify words from their first few letters (ele______) or from a fragmented form (e _ e _ h _ _t) (e.g., Graf & Mandler, 1984; Roediger, Weldon, Stadler, & Riegler, 1992). Priming on such tests seems to depend on accessing a perceptual record of experience (Kirsner & Dunn, 1985) or access to a lexical representation (Weldon, 1991), and manipulation of types of orienting task appears not to alter the basic perceptual record or lexical activation. The implication is that even the shallowest graphemic levels of processing (case judgments, checking for e's) seems to permit such processing that is needed to support priming on perceptual implicit memory tests. If graphemic processing arrested perception at a very early level (e.g., access of only visual features, but not of word forms), as postulated by the original levels-of-processing framework, then no priming might be expected to occur from the shallowest levels of processing.

In the wisdom of hindsight, the effect of orienting tasks (the "levels-of-processing effect") that seemed so powerful on certain tests does not have wide generalizability. Many tests (especially certain implicit tests) that validly assess retention show no levels of processing effect. This outcome fits with ideas about transfer-appropriate processing, but not with the original levels-of-processing framework. As noted above, the levels-of-processing framework that emphasizes meaning as a key element of mental representations is important, but has the unfortunate consequence of de-emphasizing other dimensions that in the right circumstances are themselves critical to memory performance (Kolers & Roediger, 1984). In fairness, the original levels-of-processing framework was designed to account for performance on what are today called explicit memory tests and one could say that implicit tests lie outside the boundary conditions of the framework. This is true, but the bulk of this chapter has pointed out problems even in the domain of explicit tests, and, of course, the levels-of-processing effect is not obtained on certain explicit tests (rhyme recognition, rhyme cued recall, recognition of graphemic differences; Fisher & Craik, 1977; Jacoby, 1975; McDaniel et al., 1978; Morris, et al., 1977; Stein, 1978). The transfer-appropriate processing framework provides a superior account of most of the findings problematic for the levels-of-processing framework, but in fairness some puzzles pointed out earlier in this chapter and elsewhere (e.g., McDermott & Roediger, 1996) do not fit well within that framework, either.

☐ Conclusions

The original levels-of-processing framework proposed that items could be processed to various levels of perceptual analysis and that this idea was helpful in understanding many effects in the psychology of memory. In particular, orienting tasks were thought to arrest processing at various levels of perceptual analysis (graphemic, phonemic, and semantic for verbal materials) and retention was said to be a function of the depth of perceptual analysis. The literal version of this theory was recognized as inadequate to handle the data, if not downright wrong, almost as soon as it was proposed. Even in the 1972 paper, Craik and Lockhart wrote that "'spread' of encoding might be a more accurate description" than levels of processing [for various reasons] but that "the term 'depth' will be retained as it conveys the flavor of our argument" (p. 676). In later papers, Craik and his colleagues distanced themselves even further from the strict levels-of-processing ideas, preferring other descriptors to capture encoding differences, such as elaboration, spread, distinctiveness, richness, and the like (e.g., Craik & Tulving, 1975; Lockhart, Craik, & Jacoby, 1976; Moscovitch & Craik, 1976). However, none of these terms or ideas has been specified well in the intervening years and none has received general agreement in the field as a replacement for the terms "depth" or "levels" of processing, which still reigns supreme.

Few of the points made in this chapter are really new or original. In fact, Craik and Lockhart and their collaborators made most of them in the 1970s, so a chapter like this one could have been written in 1980. We have known since at least then that the levels-of-processing framework—the original version that emphasizes strict levels of perceptual processing—is just wrong. Curiously, it was originally criticized as being circular and untestable by several commentators (e.g., Baddeley, 1978; Eysenck, 1978; and T. O. Nelson, 1977), but Craik and his collaborators provided many telling observations that called the framework into question. Their research conducted on the so-called levels-of-processing paradigm (as embodied in, say, the Craik and Tulving, 1975, experiments) undercuts rather than supports the levels-of-processing framework for the reasons given above and for others (e.g., D. L. Nelson, 1979). Hence the paradigm is misnamed and might be called (more neutrally) the Craik–Tulving paradigm. Today the term *levels of processing* is used more generically by the field, to emphasize the fact that different types of processing of words during encoding provide various levels of memory performance (albeit depending on the test used to assess retention).

The strict levels-of- (perceptual) processing framework may be wrong, but in our opinion the contributions from this approach still represent one of the most valuable developments in the psychology of memory in the 115-year history of its empirical study. Craik and Lockhart's (1972) paper deserves all of its fame, for many reasons. First, it challenged the existing conception of memory as composed of various stores. This has always been the received wisdom about memory—people are thought to search through various stores to locate memories (Roediger, 1980). Second, it began a new way of thinking about remembering, not as static traces in the head, but rather as actions and activities. The critical focus was on processing. Processing approaches to remembering, as embodied in the work of Kolers (1979) and others, still have not been satisfactorily developed, but to the extent that processing approaches have developed at all, the credit goes to Craik and Lockhart for blazing the trail. Third, the levels-of-processing framework has been fruitful and generative, providing a powerful set of experimental techniques for exploring the phenomena of memory. The manipulation of various orienting tasks predated the 1972 paper by many years and was used as an experimental technique sporadically, to address various issues. However, since the 1972 paper, the manipulation of orienting tasks has taken center stage and represents one of the most frequently used techniques in the experimental study of memory.

In sum, the levels-of-processing framework has been of unsurpassed value in the psychology of memory. The purpose of this chapter is to point out that the original framework emphasizing levels of perceptual processing cannot explain the basic levels of processing effect (or the effect of orienting tasks on recall and recognition). The field has ignored this conclusion, despite the fact that Craik and Tulving (1975) and many others pointed it out in no uncertain terms a quarter of a century ago. We may hope that in the next 30 years, unlike the last 30, researchers and theorists will attempt to correct this situation and come to grips with these phenomena both theoretically and empirically.

Acknowledgments

We thank Fergus Craik, Morris Moscovitch, Robert Lockhart, Endel Tulving, and Michael Watkins for their helpful comments on an earlier draft of this chapter. The writing of this chapter was supported by a contract from the Office of Technical Services of the U.S. Government awarded to the first author.

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 385–403). Hillsdale, NJ: Erlbaum.

Arbuckle, T. Y., & Katz, W. A. (1976). Structure of memory traces following semantic and nonsemantic orienting tasks in incidental learning. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 362–369.

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

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 657–668.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 331–354). Hillsdale, NJ: Erlbaum.

Cermak, G., Schnorr, J, Buschke, H., & Atkinson, R. C. (1970). Recognition memory as influenced by differential attention to semantic and acoustic properties of words. *Psychonomic Science, 19,* 79–81.

Challis, B.H., Velichkovsky, B. M., & Craik, F.I.M. (1996). Levels-of-processing effects on a variety of memory tasks: New findings and theoretical implications. *Consciousness and Cognition, 5,* 142–164.

Chow, P. C. P., Currie, J. L., & Craik, F. I. M. (1978). Intentional learning and retention of words following various orienting tasks. *Bulletin of the Psychonomic Society, 12,* 109–112.

Craik, F. I. M. (1973). A "levels of analysis" view of memory. In P. Pliner, L. Krames, & T. M. Alloway (Eds.), *Communication and affect: Language and thought* (pp. 45–65). New York: Academic Press.

Craik, F. I. M. (1977). Depth of processing in recall and recognition. In S. Dornic & P. M. A. Rabbitt (Eds.), *Attention and performance VI* (pp. 679–697). Hillsdale, NJ: Erlbaum.

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

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Eysenck, M. W. (1978). Levels of processing: A critique. *British Journal of Psychology, 69,* 157–169.

Fisher, R. P., & Craik, F. I. M. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 701–711.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23,* 553–568.

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

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

Jacoby, L. L. (1975). Physical features vs. meaning: A difference in decay? *Memory & Cognition, 3,* 247–251.

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

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Jenkins, J. J. (1974). Can we have a meaningful theory of memory? In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 1–20). Potomac, MD: Erlbaum.

Kirsner, K., & Dunn, J. C. (1985). The perceptual record: A common factor in repetition priming and attribute retention. In M. I. Posner & O. S. M. Martin (Eds.), *Mechanisms of attention: Attention and performance, XI.* Hillsdale, NJ: Erlbaum.

Kolers, P. A. (1979). A pattern-analyzing basis of recognition. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 363–384). Hillsdale, NJ: Erlbaum.

Kolers, P. A., & Roediger, H. L., III. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23,* 425–449.

Lockhart, R. S., & Craik, F. I. M. (1978). Levels of processing: A reply to Eysenck. *British Journal of Psychology, 69,* 171–175.

Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology, 44,* 87–112.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. L. (1976). Depth of processing, recognition, and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 75–102). London: Academic Press.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

Mathews, R.C. (1977). Semantic judgments as encoding operations: The effects of attention to particular semantic categories on the usefulness of interitem relations in recall. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 160–173.

McDaniel, M. A., Friedman, A., & Bourne, L. E. (1978). Remembering the levels of information in words. *Memory & Cognition, 6,* 156–164.

McDermott, K. B., & Roediger, H. L. (1996). Exact and conceptual repetition dissociate conceptual memory tests: Problems for transfer appropriate processing theory. *Canadian Journal of Experimental Psychology, 50,* 57–71.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Moeser, S. D. (1983). Levels of processing: Qualitative differences or task-demand differences? *Memory & Cognition, 11,* 316–323.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior, 15,* 447–458.

Nelson, D. L. (1979). Remembering pictures and words: Appearance, significance, and name. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory,* (pp. 45–76). Hillsdale, NJ: Erlbaum.

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

Packman, J. L., & Battig, W. F. (1978). Effects of different kinds of semantic processing on memory for words. *Memory & Cognition, 6,* 502–508.

Postman, L., Adams, P. A., & Phillips, L. W. (1955). Studies in incidental learning. II. The effects of association value and of the method of testing. *Journal of Experimental Psychology, 49,* 1–10.

Postman, L., & Kruesi, E. (1977). The influence of orienting tasks on the encoding and recall of words. *Journal of Verbal Learning and Verbal Behavior, 16,* 353–369.

Roediger, H. L., III. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition, 8,* 231–246.

Roediger, H. L., III, Payne, D. G., Gillespie, G. L., & Lean, D. S. (1982). Hypermnesia as determined by level of recall. *Journal of Verbal Learning and Verbal Behavior, 21,* 635–655.

Roediger, H. L., III, & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, 6,* 296–305.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.

Roediger, H. L., III, Weldon, M. S., Stadler, M. L., & Riegler, G. L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 1251–1269.

Seamon, J. G., & Virostek, S. (1978). Memory performance and subject-defined depth of processing. *Memory & Cognition, 6,* 283–287.

Stein, B. S. (1978). Depth of processing reexamined: The effects of precision of encoding and test appropriateness. *Journal of Verbal Learning and Verbal Behavior, 17,* 165–174.

Tresselt, M. E., & Mayzner, M. S. (1960). A study of incidental learning. *The Journal of Psychology, 50,* 339–347.

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

Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62,* 74–82.

Tulving, E. (1979). Relation between encoding specificity and levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels-of-processing in human memory* (pp. 405–428). Hillsdale, NJ: Erlbaum.

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

Weldon, M. S. (1991). Mechanisms underlying priming on perceptual tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17,* 526–541.