list recalled 59% of the paired associates. When tested in the other room, however, their recall dropped to 46%.

We should be somewhat cautious in interpreting these findings, because further research by Glencoe indicates that, although context can be a significant factor in retention, it can be mitigated by other retrieval cues. In short, the situation is more complex than originally thought. The study nevertheless makes an important point: If the subjects had been graded on their performance, they drop-off would have meant the difference between an A and a B, which leads to another necessary and practical point: If teachers are serious about getting optimal performance from their students, then tests and finals should always be given in the same classroom in which the class meetings took place.

At or near levels of Processing

As you'll recall, by the end of Chapter 3 we had seen that cognitive psychologists were becoming convinced that complete understanding of memory would not result from specifying the characteristics of the separate memory buffers. Indeed, cognition was instead trying to understand the many ways in which it might be possible to encode a stimulus. Craik and Lockhart (1972) were among the earliest researchers to express this viewpoint. They rejected the idea that memory's location determined its characteristics. Rather than think of the stimulus to be remembered as a fixed object with distinct properties that were altered as it moved through a rigid system of storages and buffers, Craik and Lockhart maintained that the stimulus could be processed in various ways. An individual could bring sensory processes to bear on a stimulus and extract from it its physical characteristics. On the other hand, the person had control over other cognitive processes capable of extracting and encoding acoustic or semantic features of the stimulus. Craik and Lockhart viewed this as a continuum of progressively deeper cognitive processing—deeper in the sense that more background knowledge is required to carry out a semantic analysis of a word than an acoustic analysis. Material stored in memory—the memory code—acquired its semantic or acoustic properties not because it was being stored in specific locations but because it had been processed in ways that were under the person's control. The memory code is therefore a record of the cognitive processes that have been performed on it. In summary, each approach to processing produced a cognitive code that could be evaluated along a continuum of depth. Generally, the greater the semantic analysis (the more meaning extracted from the stimulus), the greater the depth of processing.

Craik (1979) stated that the levels-of-processing model has two main postulates. First, semantic analysis results in a deeper and thus more meaningful code than does acoustic analysis. Second, the deeper the code, the more durable the memory. Thus forgetting is simply a function of depth of processing: we forget things that we have not processed semantically. To these postulates I add corollaries. First, be aware that we're not dealing with a multiple-storage model. That is, we have no notion about transfer of information from one storage location to another. Second, and correspondingly, we see no capacity limitations. Evidence that seems to indicate the need for different storages with differing capacities (such as the memory-span phenomenon) can be interpreted as processing limitations. Another implication is that durability of a memory is somewhat independent of the time spent processing. A great deal of time spent processing material at a nonsemantic (shallow) level probably will not produce a more durable memory than would a short time processing at the semantic level (Craik & Watkins, 1973; Randus, 1977).

A study by Parket (1984) illustrates the depth-of-processing effect. His subjects were given a word, about which they had to make a semantic-orienting or nonsemantic-orienting decision. The orienting decision refers to a judgment about the word. For example, a semantic-orienting decision might involve a category or synonym judgment. A nonsemantic task might involve making a judgment about how many vowels the word had or whether it had been printed completely in capital letters. Following a series of such trials, the subjects were given a surprise free-recall test. Those who had been semantically oriented recalled significantly more of the target words than did those who were nonsemantically oriented. This result suggests that semantic processing produces a more durable memory code than does nonsemantic processing.

A study by Jacoby, Craik, and Begg (1979) supports this interpretation. Their subjects were given pairs of common nouns (e.g., house-boot) and were told to evaluate the difference in size between the objects on a 1 (not much difference) to 10 (very different) scale. The difference between some of the named objects was relatively small, whereas for others the difference was large. After the subjects had made these evaluations, they were given an unexpected memory test in which they were asked to recall as many of the objects as they could. Jacoby et al. found an inverse relationship between the size difference between the objects and the likelihood of their recall. The subjects were more likely to recall the objects when the difference between them was small. Jacoby et al. explained that the task required a semantic analysis of the objects' properties. When the objects were approximately the same size, however, deeper analysis was required, producing a more durable memory code. Notice that the subjects were not aware that a memory test was forthcoming when they evaluated size differences. But their being engaged in a semantic analysis of the words facilitated their retrieval. The implications are clear: Your choices of retrieving a memory depend on the type of processing you do to remember it, not on how hard you try to remember it. These effects were shown in a study by Hyde and Jenkins (1973). The subjects saw a list of 24 words, which was presented at the rate of 3 seconds per word. Each subject had one of two tasks. In one condition subjects simply had to check whether each word had an a or an e. In the second, they had to rate the pleasantness of the word, which presumably required deeper processing than simply looking at the word's physical characteristics. This study had a second variable. Half the group of subjects, the intentional learning group, was told that a memory test would be given after the words were presented, and they were encouraged to learn the words. The other half of the group were not informed about the subsequent memory test, and any knowledge they retained about the words was therefore incidental.

Table 4.3 shows the findings in this study. Notice the strong depth-of-processing effect. When the subjects carried out the semantic analysis, their recall was dramatically enhanced. Notice, too, that incomprehensibility had little effect on the proportion of the words recalled. The levels-of-processing viewpoint thus accounts for something
that vexed multiple-storage models. Our memories are not controlled directly by our intentions (which makes the Atkinson and Shiffrin control process somewhat misleading) but rather by the type of processing we do.

**Maintenance Rehearsal and Elaborative Rehearsal**

To help make these processing effects clearer, Craik and Lockhart (1972) distinguished between two kinds of rehearsal. Type I, sometimes called maintenance rehearsal, refers to the continual repetition of analyzers that have already been carried out. It does not lead to stronger or more lasting memories, its principal function is to retain the availability of an item in memory. Maintenance rehearsal tends to emphasize the phonetic aspects of a stimulus (Wickens, 1984). Type II rehearsal, also known as elaborative rehearsal, refers to successively deeper processing of the stimulus and does produce more durable memories. It emphasizes the stimulus’s semantic aspects. Both types are under the person’s control: they operate on material in working memory, and they compete with other tasks for cognitive resources.

Craik and Watkins (1973) demonstrated that the kind of rehearsal, not how much, determines a memory’s durability. The subjects in their study were presented with twelve lists, each with twelve words. The subjects were told to consciously rehearse these words. Moreover, they were told that the last four words in each list were especially important and that these should be remembered at all costs. To emphasize this instruction, the last four words in each list were printed in bold letters. The study had two retrieval conditions. Subjects were asked to recall the list either immediately or following a 20-second delay. If recall was delayed, the subjects were told that actively rehearsing the list was permitted. When all twelve lists had been presented, Craik and Watkins sprung a surprise: they asked the subjects to recall as many of the entire set of 144 words as they could. The findings are shown in Figure 4.1.

As the top graph shows, the subjects took advantage of the opportunity to rehearse the last four items in the lists during the 20-second delay. As the lower-left graph shows, an expected recency effect occurred when the words were recalled immediately after presentation or after the 20-second delay. The lower-right graph shows the effects of the surprise. Even though the last four items in each list were rehearsed far more frequently than the others, and even though an immediate recency effect occurred, virtually no long-term effect appeared. In the final recall test, words that had been presented in one of the last four serial positions were not recalled with any greater frequency than words presented in other serial positions. Even though the last four words had been rehearsed more than the others, this was maintenance rehearsal, which did not necessarily lead to more durable memories (Craik & Watkins, 1973).

![Figure 4.1](image)

The distinction between elaborative and maintenance rehearsal sounds plausible, and the Craik and Watkins results make sense. This position implies, however, that all semantic processing is equal and thus should create memories of equal durability. But does it?

To answer this question, Craik and Tulving (1975) presented sentences and words tachistoscopically to their subjects. The words and sentences were shown consecutively, and the subject’s task was to decide whether the word would meaningfully fit into a blank left in the sentence. The sentence’s semantic complexity was varied, too. These three examples show the increasing complexity:

**Simple:** She looked at the cat.
**Medium:** The cat frightened the children.
**Complex:** The great bird swooped down and carried off the struggling boy.

After sixty judgments (twenty at each complexity level), Craik and Tulving sprang the by-now-familiar surprise: a memory test, this time in the cue-recall format. The subjects were given the sentence and asked to recall the word that had been shown consecutively. The surprise now was on Craik and Tulving: subjects remembered more of the complex-sentence fill-ins than the simple-sentence fill-ins. Considering only yes responses, those in which the word could be meaningfully used in the sentence, the...
subjects recalled about twice as many of the complex sentences, but only for an answer. Semantic processing had been carried out, which is the deepest level according to the theory. Therefore, no differences should have appeared in retrieval as a function of the sentence's complexity. Why, then, did Craik and Tulving observe these differences? The answer must lie in the elaborative rehearsal and the larger cognitive structure into which the elaborated code was being fed. In other words, the semantic processing done by subjects on the complex sentences seems to have accessed other cognitive codes richer or more elaborate than those accessed in the simple sentences.

Some Problems with "Levels of Processing." Although the levels-of-processing account is persuasive, it has been dogged by a few problems. First, the approach has the problem we've just touched on—contradictory findings. Nelson and McEvoy (1979) reasoned that, if words have been processed at the semantic level, then providing a semantic cue should produce better retrieval than a nonsemantic cue. But as it turns out, presenting nonsemantic cues (such as the cue "TEAT" for the list word DIME) is just as effective as presenting semantic cues (such as "an American cow" for DIME). Further, Hunt and Elliot (1980) demonstrated that words with irregular and distinctive orthography (such as phlegm) are retained better when they are processed as part of a list with words in both regular and irregular orthography than when a whole list is made up of such distinctive words. Even when the task requires semantic analysis, some nonsemantic information—this tinct orthographic distinctiveness—seems to be retained.

A second problem with this approach is an independent definition for depth of processing. Usually, the level of processing is operationally defined by the orienting task the subject is given to do. For example, if you are given a task involving checking letters of a rhyme, this processing is considered nonsemantic. But if you are required to produce a synonym, then you must be processing the material at the semantic level. Many commentators have pointed out (Nelson, 1979; Postman, Thorndyke, & Gray, 1978) that this operational definition is not quite adequate. It's hard to say exactly what the subject is doing when instructed to process the material in a particular way. Furthermore, it's hard to see a way around the problem of defining depth. Linking depth of processing to time spent processing is irrelevant by definition, and asking the subjects to self-report their own depth of processing is fraught with difficulty (Scannon & Virostek, 1978).

Yet another problem is the relationship between depth of processing and automaticity. As we saw earlier, highly practiced tasks become progressively more automatic—that is, they can be executed without heavy demands on cognitive processes. What if subjects become highly practiced at naming semantic decisions? According to the levels-of-processing view, the subject should show good retention of this material because it has been deeply processed. According to the automaticity viewpoint, though, the subject should show little or no retention of such material. Fiske and Schneider (1984) pitted these viewpoints against each other. Their subjects were extensively trained in categorizer materials automatically. Categorization is an orienting task that is usually thought of as semantic. But Fiske and Schneider found that their subjects, who had done well at the categorizer task, showed little recognition

memory for the categorizer materials. This finding challenges the levels-of-processing approach. A problem can also be found in the maintenance and elaborative real concept. At least in some tasks, maintenance rehearsal does improve memory (Glenberg & Adams, 1978). That kind of finding suggests to many students of memory that a categorical view of rehearsal strategies is probably not correct (Craik, 1979; Jacoby & Craik, 1979). Rehearsal strategies can probably be graded along a continuum of elaboration. This formulation will almost certainly be advanced in future work.

Transfer-Appropriate Processing. Perhaps the biggest complication for the levels-of-processing position comes from a classic study done by Morris, Gradmann, and Prakas (1977). In this study the experimenter read aloud thirty-two sentences, each missing one word. In the shallow-processing condition, the experimenter might read a sentence such as this: "Blank rhymes with legal." After hearing this sentence, the subjects heard an additional word, the target word, and their task was to decide if the target word could be substituted appropriately for the blank. The subjects should say yes if the target word was "legal" but no if it was "poach." In the deep-processing condition, the task was similar, but was modified slightly to demand semantic processing. Now, the experimenter might read a sentence such as: "The blank has a silver edge." Hearing the target word "edge" after this sentence should produce a no from the subjects, whereas hearing the target word "train" should be answered yes. At test time, half the group of subjects were given a standard recognition task: each of the target words was presented along with additional words that functioned as distractors. The subjects were to pick out the target word. Consistent with the levels-of-processing position, the subjects did a better job recognizing target words that had been presented in the semantic version of the task than those presented in the rhyming version. But a different fate awaited the other half of the group. These subjects were given a rhyming recognition task. They got a series of words and were asked to pick out the word that rhymed with a target word that had been seen before. Continuing the example above, if the original target word had been "legal," the subject might see a series with the word "legal," and of the original target word had been "train," the subject might now see a series including "train." Now, something surprising happened: The subjects did a better job of picking out rhyming words of targets originally presented in the rhyming condition than they did for rhyming words of targets presented in the semantic condition. This finding is exactly the opposite of that predicted by the levels-of-processing position; here, when subjects engaged in the shallow, rhyming task, they did better than when they had engaged in the deep, semantic task. How come?

Morris et al. used transfer-appropriate processing to explain these effects: the cognitive processes that are used in the initial learning or encoding of some material interact with the cognition used at retrieval time. Thus the best encoding is that based on cognitive processes that most closely match the type of cognitive processing that will be used at retrieval time. Simply, if you have to retrieve something by rhyming, then the best encoding you can make also will involve rhyming. But if you have to retrieve material using semantic processing, then it is probably best for you to use semantic processing at encoding time.
Intrachapter Summary and Interpretation

We've been exploring an alternative to the storage position that we examined in Chapter 3. Although this alternative isn't necessarily explained by an all-encompassing metaphor such as "storage," we might think of it as something like a "recomputation" metaphor. That is, from this perspective, rather than think of the memory as a cognitive device that can store and retrieve things, we might see it as a device for "recomputing" an experience, using the stimulation that we're currently experiencing. As we've seen, this perspective provides us with a good basis for understanding some of the phenomena that Bartlett discovered so many decades ago. Why would students retrieve incorrect elements from the folk tale months or years later? Although it could be true that the students had simply stored and retrieved the wrong information, it might be more congruent to explain their errors as caused by faulty reconstruction of the facts. And it also seems likely that such errors would be at the time of retrieval, rather than at the time of initial storage. From these we went on to consider how context possibly "biased" the memory system to create constructions and therefore pathways toward reconstructing specific memories. We haven't completed our discussion yet, but as we have been, the type of cognitive processing that you do to remember something strongly biases the likelihood of retrieving (or perhaps I should say, of reconstructing) that memory at retrieval time.

Encoding Specificity

We've been considering the effects of various contextual variables on the likelihood of our retrieving material that we've tried to learn. Beyond the effects of various external stimuli, we've seen that the cognitive processes themselves that we use to learn and to retrieve influence the likelihood of retrieving a specific memory. If you're a student, this finding probably makes sense.

For example, as a student, you're subjected to various tests. Some of these probably are essay tests or oral exams, along with true-false or multiple-choice questions. Essay and oral exams measure your ability to recall material from memory, and multiple-choice questions tap your ability to recognize stored material. The distinction between recall and recognition often boils down to the number of cues or prompts provided. And as you may be aware, knowing which kind of test you have to face strongly influences the way in which you study. That is, knowledge of the retrieval task influences the encoding you do.

This influence was well demonstrated in a study by Leonard and Whitten (1983). Half their subjects studied a list of words with the expectation that they would be given a recognition task. That task would be arranged like a multiple-choice test; the subjects would be required to pick out the listed word from among several alternatives. The other subjects studied the list with the idea that they would be given a free-recall test. But these subjects were deceived. All the subjects were tested with the multiple-choice procedure. On some of the items, the word previously studied was presented in the context of semantically related words. On other items, the studied word appeared in the context of semantically unrelated words. Leonard and Whitten found that the subjects who had studied for recognition showed a decrement in performance when the words were presented in a semantically related context, but this effect was not observed in the subjects who had studied for recall.

This finding makes sense when you think about the task from the subject's point of view. Suppose one of the words in the list was evil. Subjects who expected a multiple-choice test may have prepared themselves to pick out a word that means "bad," "nasty," or "rotten." Subjects who expected a free-recall test had to do something different. They had to establish some sort of context for evil that they could reproduce from scratch at retrieval time. Consequently, when the alternatives were semantically related to the word studied, the subjects who had expected a recognition task apparently had trouble deciding which was the target word and which was the context.

The Llouard and Whitten (1983) findings have several implications. First, they imply that subjects use different retrieval strategies depending on the memory task. Second, the findings suggest that the subjects know ahead of time that they will use different strategies to remember the material depending on the task. A more thorough discussion of this phenomenon awaits us in later chapters. Third, the findings suggest that the intended retrieval strategy affects the actual encoding.

Retrieval processes in recognition and recall have been thought to differ in this way. In a recall test with a word list, the subject generates candidate items and then decides about each candidate's inclusion or exclusion from the list studied (Hulse, Dese, & Egeth, 1975). In a recognition task, this sort of procedure need not take place, because the candidates have already been provided by the experimenter. According to this view of retrieval, performance on recognition tests should always be superior to performance on recall tasks. To recall something, we have to do two things: generate the candidate and then recognize that it belongs on the list. But to recognize something, we have to do only one thing: recognize it. The generation stage can be bypassed (Wescott, 1982).

When Recall Beats Recognition: A striking series of studies (Flesner & Tulving, 1978; Tulving & Thompson, 1973; Watkins, 1974; Watkins & Tulving, 1975) demonstrates that under some conditions, recall is superior to recognition, and that subjects are sometimes able to recall material that they cannot recognize. Watkins (1974) gave his subjects lists of paired-associate nonsense words. Each pair had a five-letter A part and a two-letter B part. Although the A and B parts were not meaningful by themselves, the combined seven-letter item was as if it were five parts (invol-v-E). After one presentation of the list, recognition memory was assessed by giving the subjects a list of the A parts they had just seen against the context of other two-letter nonsense syllables. In a later assessment of cued-recall memory, the subjects were prompted with the A part and had to recall the B part of the pair. Recognition accuracy was a dismal 9%, but cued-recall accuracy was 67%.

This effect was extended in a study by Watkins and Tulving (1975). Their subjects were given a list of paired associates such as HEAD-LIGHT, but they were told that they would be responsible only for the second word in the pair, which was called the "to-be-remembered" word, or TBR. After the subjects had studied the list, they were given a word and were asked to generate the first four free associates that entered their minds. For example, the subject might be given the word dark, and the associates light, night, shadow, and pitch might be produced. The words given to the subjects