listened to a tape recording of the 24 words, they tried to recall as many as they could, in any order.

The other three groups were incidental learning groups who were not informed that they should try to remember the words. They heard the same recording of 24 words but were asked to make a judgment about each item on the list. One group simply rated the words as pleasant or unpleasant, another group judged whether each word contained the letter e, and a third group estimated the number of letters in each word. The purpose of using these orienting tasks was to try to create different levels of processing. The first group would have to consider the meaning of the words. The latter two groups would have to consider the spelling of the words, the meaning of the words would be irrelevant to them. Since, according to the levels-of-processing theory, semantic processing should result in better recall than nonsemantic processing, the undergraduates who rated the pleasantness of the words should show better recall than those who considered the spelling of the words.

The results supported the prediction. The average number of words recalled was 16.3 for those students who rated pleasantness, 9.9 for those who estimated the number of letters, and 8.4 for those who judged the presence of the letter e. The most striking aspect of the results is that students in the pleasant-unpleasant group recalled virtually as many words as those who were told to try to learn the words (16.3 versus 16.1). In other words, incidental learning was as effective as intentional learning when the students considered the meaning of the words.

We have been assuming, along with Hyde and Jenkins, that differences in recall among the three incidental groups were caused by the possibility that the students in the pleasant-unpleasant group were more likely to attend to the meaning of the words than the students in the other two groups. Do we have any direct evidence for this assumption? The fact that the list consisted of pairs of words that are semantically related provides a clue. Recognizing that words are related in meaning can make it easier to recall them. For example, the recall of green may remind a person that red was also on the list. One indication that people were attending to the meaning of the words would be if they recalled the primary associates together—red followed by green or vice versa.

Hyde and Jenkins defined the percentage of clustering as the number of associated pairs recalled together, divided by the total number of words recalled. The amount of clustering was 26% for the group that made judgments about the letter e, 31% for the group that estimated the number of letters, 68% for the group that was told to study the words, and 68% for the group that judged the pleasantness of the words. These results support the assumption that the groups differed in how much they used meaning to aid recall. Those groups that were the most sensitive to the meaning recalled the most words.

Table 6.2 Typical questions used in levels-of-processing studies

<table>
<thead>
<tr>
<th>Level of Processing</th>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural</td>
<td>Is the word in capital letters?</td>
<td>TABLE</td>
<td></td>
</tr>
<tr>
<td>Phonemic</td>
<td>Does the word rhyme with WEIGHT?</td>
<td>MARKET</td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>&quot;He met a ___________ in the street?&quot;</td>
<td>FRIEND</td>
<td>closed</td>
</tr>
</tbody>
</table>

about a word's meaning resulted in better memory than those about a word's sound or the physical characteristics of its letters.

The left half of Figure 6.2 shows the average response time required to answer the three kinds of questions. The case questions could be answered most quickly, followed by the rhyme questions, followed by the sentence questions. Although the recognition results on the right can be predicted from the response times on the left, it is not always true that slower responses lead to better memory. It is possible to design a structural decision task that results in slow responses and poor retention.

Imagine that an experimenter shows you a card with a five-letter word such as stop or black. Your task is to respond positively if the word consists of two consonants followed by two vowels followed by a consonant and negatively for any other sequence of consonants and vowels. As you might guess, your response times would be relatively slow. In fact, it takes about twice as long to make this kind of structural decision as to make a semantic decision about whether a word fits into a sentence. If good retention is caused by long response times, the structural processing should now result in better retention than the semantic processing. However, recognition is still much better after semantic processing, proving that the level of processing, not the time spent processing, is the best determinant of retention.

**CRITICISMS AND MODIFICATIONS OF THE THEORY**

**Criticisms**

The levels-of-processing theory has had a major impact on memory research—many investigators designed studies to explicitly test its implications; others found it a convenient framework in which to discuss their results. Because much of this research was quite supportive of the theory, it wasn't until about 5 years after the Craik and Lockhart paper that psychologists began to seriously question the usefulness of the theory (Nelson, 1977; Baddeley, 1978; Eysenck, 1978). One of the main criticisms was that it was too easy to account for differential rates of forgetting by appealing to the theory. An investigator might claim that differences in rates of forgetting were caused by differences in levels of processing, without measuring the levels of processing.

To avoid this criticism, it is necessary to be able to measure depth of processing independently of retention. The argument that depth increases from structural to phonemic to semantic processing appealed to most psychologists because it is consistent with the ordering of the information-processing stages shown in Figure 1.1 (page 5). Analyzing the physical structure of a pattern leads to retrieving its name, which in turn lends to considering its meaning by retrieving stored associations from LTM. One problem with this assumption is that, although this sequence provides a reasonable account of how information is analysed, it is not a necessary sequence (Baddeley, 1978; Craik, 1979a). Although Craik and Lockhart originally hoped that encoding time would provide an independent measure of depth of processing, we have seen that this measure has its limitations (Craik & Tulving, 1975).

Another difficulty with the concept of depth of processing is that, even if we had an objective ordering of the "depths" of different memory codes, it still would not tell us why some codes are more effective than others. Why are semantic codes better than phonemic codes and phonemic codes better than structural codes? Psychologists have suggested two possible answers. One is that memory codes differ in elaboration they are, and more elaborate codes result in better memory. The other is that memory codes differ in distinctiveness, and more distinctive codes result in better memory.

**Elaboration of Memory Codes**

One explanation of how memory codes differ proposes that they differ in the number and types of elaborations stored in memory (J. R. Anderson & Reder, 1979). This view assumes that people store much more than simply the items presented to them—they also store additional associations that help them remember the items. Anderson and Reder have proposed that, although it is very easy to elaborate material at the semantic
level, it is difficult to construct elaborations at the structural or phonemic level. Most of the associations we have are concerned with meaning rather than with the physical structure of letters, spelling, or pronunciation. Anderson and Reder suggest that the reason for this difference is that people usually try to remember the meaning of what they read rather than such details as what the letters looked like. As a consequence, people have learned to elaborate on the semantic content because doing so is generally more useful than elaborating on nonsemantic content.

One virtue of the elaboration hypothesis is that it provides a possible explanation of how differences can occur within a particular level of processing (Craik, 1979b). Although the original levels-of-processing proposal predicted that semantic processing should be superior to nonsemantic processing, it could not account for differences in retention for the two different semantic tasks. The elaboration hypothesis predicts that such differences should occur if the two tasks differ in the extent of semantic elaboration.

One method for increasing semantic elaboration is to provide a richer, more elaborate context. This approach is illustrated by one of the experiments in the Craik and Tulving (1975) study. The experiment tested for the recall of words after a semantic judgment task in which people determined whether a word would fit into a sentence frame. There were three levels of sentence complexity—simple, medium, and complex. For example:

Simple: She cooked the _______.
Medium: The ripe _______ tasted delicious.
Complex: The small lady angrily picked up the red _______.

After completing 60 judgments, subjects were asked to recall as many words as they could from the initial phase of the experiment. They were then shown the original sentence frames and asked to recall the word associated with each sentence. The first part of the recall task is called noncued recall, and the second part is called cued recall because subjects could use the sentence frames as retrieval cues. Figure 6.3 shows the proportion of words recalled as a function of sentence complexity. Sentence complexity had a significant effect on recalling words that did fit the sentence. This was true for both cued recall (CR—yes) and noncued recall (NCR—yes), although the effect was greater for cued recall. The effect of sentence complexity supported Craik and Tulving's hypothesis that more complex sentence frames would produce a more elaborate memory code and would improve recall.

The more elaborate code was ineffective, however, if the word did not fit the sentence. This finding suggests that the elaboration must be consistent with the meaning of the word in order to be effective. Even when elaboration is generally consistent with the meaning of a word, it can vary in effectiveness depending on how precisely it relates to the words meaning. Imagine that you read the sentence "The fat man read the sign." Some time later someone shows you the same sentence with the word fat replaced by a blank and asks you to recall the missing word. If elaboration is effective, you might do better if you read an elaborated sentence such as:

1. The fat man read the sign that was 2 feet high, or
2. The fat man read the sign warning about thin ice.

Although both sentences provide additional information, there is an important distinction between the two elaborations. The first is an imprecise elaboration because there is no apparent relation between the adjective fat and the height of the sign. The second is a precise elaboration because the degree of danger of thin ice depends on a person's weight.

B. S. Stein and J. D. Bransford (1979) tested the effectiveness of precise and imprecise elaboration by comparing two groups of students in an incidental learning task. Students in the control group read ten short sentences and were told that the purpose of the experiment was to measure sentence comprehensibility. The second and third groups of students were told the same ten sentences elaborated by an additional phrase that was either precisely or imprecisely related to a

Figure 6.3 Proportion of words recalled as a function of sentence complexity. CR = cued recall; NCR = noncued recall

target word in the sentence. A fourth group of students were told to generate their own elaborations to the sentences, which could measure the probability that certain phrases would be generated. At the end of the experiment, everyone was shown the unelaborated sentences and asked to recall a missing target word.

Students in the control group recalled an average of 4.2 words, compared with 7.2 words for the precise elaboration group, 7.4 words for the precise elaboration group, and 5.8 words for the self-generation group. The results show that elaboration is not always effective in recall, since precise elaboration actually caused a decline in performance relative to the control group. To be effective, the elaboration should clarify the significance or relevance of a concept (such as 'fat man') relative to the context (skin ice) in which it occurs.

The fact that recall following self-generation was intermediate between that for precise and precise elaboration suggests that the students' elaborations contained a mixture of the two types. Two judges therefore divided the subject-generated elaborations into two groups (precise and precise), depending on whether the information clarified the relevance of the target words in the sentence. Students were able to recall 91% of the target words in the cases where they had generated precise elaborations and 49% in the cases where they had generated precise elaborations. A second experiment revealed that instructions were effective in encouraging subjects to generate precise elaborations. Subjects in the impractical elaboration group were asked to elaborate with the question "What else might happen in this context?" Subjects in the precise elaboration group were prompted to elaborate with the question "Why might this man be engaged in this particular type of activity?" Subjects in the latter group recalled significantly more target words, indicating that elaboration is particularly effective when it is directed toward understanding the potential relevance of the information presented.

Distinctiveness of Memory Codes

Memory codes can differ in distinctiveness as well as in the extent of elaboration. The term distinctiveness refers to how easy it is to distinguish one item from another. To remember something, we would like to make it really stand out from other items that could interfere with our memory. There are several different ways in which an item can be distinctive, and I will follow a classification proposed by Schmidt (1991) that distinguishes among four different kinds of distinctiveness.

One kind of distinctiveness is called primary distinctiveness in which distinctiveness is defined relative to the immediate context. Imagine that you are shown a list of common words and all the words are printed in red ink, except for one word that is printed in black ink. Later you are asked to recall the words on the list. Which word do you think you would have the best chance of recalling? The results of past research indicate that you would more likely recall the word in black ink than the words in red ink. Note that the word in black ink is distinctive only because the color differs from the color of other words on the list. In general, a common word in black ink is not particularly distinctive.

Release from proactive interference (discussed in Chapter 4) is an example of improving recall by making items distinct from other items in the immediate context. People recalled more items when the material changed from words to numbers or from numbers to words than when the material stayed the same. Recall also improved when the items changed from sports events to political events or from political events to sports events. All the changes made these items more distinct from the preceding items.

In contrast, secondary distinctiveness is defined relative to information in our LTM rather than to information in the immediate context. One example is a characteristic of a word's spelling. A word is orthographically distinctive if it has an unusual shape, as determined by the sequencing of short and tall letters in the word. Orthographically distinctive words include lymp, khaki, and afghan. Examples of orthographically common words are looky, kennel, and airway. The first three words have unusual shapes, and the last three have more typical shapes. Notice that a shape is unusual (distinctive) relative to all other words stored in LTM, not just to words in the immediate context of the experiment.

When people are asked to recall a list of words, half of which are orthographically distinctive and half of which are orthographically common, they recall significantly more of the distinctive words (R. R. Hunt & Elliott, 1980). It is clear that the shape of the words, rather than some other factor, causes the results. When the same list is presented orally, rather than visually, there is no difference in recall. There is also no difference in recall when the words are typed in capital letters; people do not recall LYMPH, KHAKI, and AFGHAN any better than LEAKY, KENNEL, and AIRWAY. Apparently, the different heights of lowercase letters contribute to the effect since all letters are the same height when capitalized.

A third kind of distinctiveness is called emotional distinctiveness and is motivated by the finding that events that produce strong emotional responses are sometimes remembered well. These events include flashbulb memories—the vivid recollections that most people have of the circumstances surrounding their discovery of a shocking piece of news (R. Brown & Kulik, 1977). Events such as the assassination of President Kennedy or the explosion of the space shuttle Challenger (Winograd & Neisser, 1992) have been studied in examples of people's flashbulb memories. I don't remember how I first learned about the Challenger, but I'm certain my emotional reaction.
brother will never forget his experience. He was driving in the direction of the Kennedy Space Center during a business trip to Orlando when he noticed that the vapor trail from the rocket's engine suddenly stopped. Suspecting that something had gone wrong, he turned on his car radio, which confirmed the failure of the launch.

Although Schmidt (1991) includes emotional distinctiveness in his taxonomy, he admits that it is not always clear which aspects of an emotional memory are enhanced or even whether the concept of "distinctiveness" provides an adequate explanation of the impact of emotion on memory. In addition, more recent evidence has questioned whether emotional events (flashbulb memories) are better remembered than ordinary events. For example, Weaver (1993) compared students' memory for an emotional event (the 1991 bombing of Iraq) and an ordinary event (a routine encounter with a roommate or friend). When their memory was tested both 3 months and 1 year later, Weaver found no difference in the accuracy of the two memories. He did find differences in confidence—students were more confident of the accuracy of their memory about the bombing but their confidence was unwarranted.

A fourth kind of distinctiveness is called processing distinctiveness. Processing distinctiveness depends on how we process the stimulus—it is therefore the result of the memory code that we create for an item rather than the characteristics of the item itself. For instance, even if an item is not very distinctive, you may think of a distinctive way of remembering it. If it is distinctive, you may think of a way of processing it to make it even more distinctive. Elaboration is one possible strategy to make an item more distinctive, but the elaboration should emphasize characteristics that differentiate that item from other items (Eysenck, 1979).

An example of processing distinctiveness is that people apparently remember faces as caricatures by exaggerating the distinctive features to make the faces even more distinct. When discussing distinctive features in Chapter 2, I gave an example of a study in which people more quickly identified line drawings of their friends when the line drawings were caricatures rather than accurate (Rhodes, Brennan, & Carey, 1987). Caricatures of unfamiliar faces are also better recognized in a standard recognition-memory test than the faces that were actually shown (Mauro & Kubovy, 1992). Undergraduates viewed 100 slides of faces constructed from an Identi-Kit, as illustrated by the two right faces in Figure 6.4. They later were shown 300 test faces and asked to indicate if each test face was exactly the same as one shown in the test series of slides. The test faces included some faces that were new, some faces that were old, and some faces that were caricatures of the old faces.

The caricatures are shown on the left side of Figure 6.4 and were created by making a distinctive feature even more distinct. The high forehead in the top face is made even higher and the long chin of the bottom face is made even longer. The interesting finding is that the caricatures were recognized significantly better than the original (old) faces. This finding is consistent with the processing distinctiveness concept—people encoded the faces into memory in a manner that made each face even more distinct than the original face.

Because both processing distinctiveness and the levels-of-processing theory emphasize the importance of creating good memory codes, it is perhaps not surprising that some psychologists have proposed that the levels-of-processing effect is caused by differences in distinctiveness. To demonstrate that distinctiveness can account for the levels-of-processing effect, it would be necessary to show that semantic codes are more distinct than phonemic codes and that phonemic codes are more distinct than physical codes. Some research has already been directed toward the first comparison. Several psychologists (Moscovitch & Gaal, 1976; Eysenck,
1979) have argued that semantic codes result in better retention than do phonemic codes because semantic codes are much more distinctive than phonemic codes. They base their argument on the fact that there is a relatively small number of phonemes; thus, phonemic codes necessarily overlap with each other, whereas the domain of possible meanings is essentially limitless.

The experimental study of elaboration and distinctiveness has modified the original conception of levels of processing (Craik, 1975). Some of the original ideas have survived, however. The central idea that there are qualitative differences in memory codes, that different orienting acts can determine which codes are emphasized, and that memory codes differ in their decay rate remains a useful conception of memory. The major shift in emphasis has been the attempt to provide a theoretical basis for these findings by determining how structural, phonemic, and semantic codes can differ in distinctiveness and elaboration.

ENCODING SPECIFICITY AND RETRIEVAL

The Encoding Specificity Principle

The change in emphasis from "levels to "elaboration" to "distinctiveness" was accompanied by another refinement in the theory. The original theory (Craik & Lockhart, 1972) had much to say about how words were coded but little to say about how they were retrieved. Yet we saw in the previous chapter that appropriate retrieval cues, such as encouraging eyewitnesses to reconstruct the context of the crime, can enhance recall. The usefulness of providing an appropriate context for facilitating retrieval is illustrated by the difference between positive and negative responses in Figure 6.2. Words that resulted in positive responses, because they either formed a rhyme or fit the context of a sentence, were recalled more often than words that resulted in negative responses. We have also seen that the use of more complex, elaborate sentence frames facilitated recall for positive responses but not for negative responses (Figure 6.3). This effect was particularly evident when the context was provided as a retrieval cue. Craik and Tulving (1975) interpreted this finding as support for their view that a more elaborate context is beneficial only when the test word is compatible with the context and forms an integrated unit. A complex sentence like "The small lady angrily picked up the red _______" makes it easier to retrieve a positive response (tomato) but does not make it easier to retrieve a negative response (table).

These results show that, under certain conditions, some retrieval cues are more effective than others. A general answer to the question of what makes a retrieval cue effective is provided by the encoding specificity principle, which has been stated as follows: "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).

Let's dissect this definition into two parts. The first part states that memory traces differ not only in their durability but also in the kind of information they contain. The second part states that the information that memory traces contain determines what kind of retrieval information should facilitate their recovery. The first part is essentially equivalent to the levels-of-processing framework; the second part forces us to take a closer look at retrieval. The second part implies that it is possible to hold constant the encoding conditions of an item and still observe large differences in its recall, depending on the retrieval conditions. The encoding and retrieval conditions can interact in the sense that a cue that is effective in one situation may or may not be effective in another.

The encoding specificity principle has usually been applied to studying how the retrieval cue relates to the memory code for the stimulus. However, the encoding and retrieval conditions can apply to a broader context such as the location in which learning occurred or even the mood of the learner. The study of mood-dependent memory tests the hypothesis that we are better able to recall information if our mood during retrieval matches our mood during learning. Although the evidence has generally been supportive of this hypothesis, the degree of support may depend on the particular paradigm used to test the hypothesis; for instance, strongest support may come from situations in which people recall information that they generated themselves. A recent study found strong support for mood-dependent memory when people had to recall autobiographical events that they had generated several days earlier (E. Eich, Macaulay, & Ryan, 1994). Subjects who were in the same (pleasant or unpleasant) mood during both encoding and retrieval recalled significantly more events than people who were in different moods. Tests of the encoding specificity principle have generally focused on the material that people have to recall rather than on where learning occurred or on the mood of the learner. We will now look at this research.

Interaction between Encoding and Retrieval Operations

A study by Thomson and Tulving (1970) provided some initial support for the encoding specificity principle. It seems intuitively obvious that the effectiveness of a retrieval cue should depend on how closely the cue is associated with the test item. The following are good retrieval cues: white for BLACK, meat for STEAK, dumb for STUPID, woman for MAN, ice for COLD, and dark for LIGHT. People who are asked to remember BLACK,