CHAPTER FIVE

Perspectives on Processing

We soon forget what we have not deeply thought about.

—Marcel Proust

The three previous chapters have emphasized memory as a structure. For example, in the modal model, one of the most important determinants of whether an item will be recalled is the memory store that holds the item: An item stored in short-term memory will be remembered differently than will an item in long-term memory or sensory memory. During the 1970s, a different view became increasingly popular—one that shifted the emphasis from the underlying store as the most important factor to the type of processing that was performed.

Levels of Processing

Craik and Lockhart (1972) offered the first detailed view suggesting that the type of processing was more important than the underlying theoretical structure. They made four key assumptions. First, they conceptualized memory as the result of a series of analyses, each at a deeper level than the previous one, that are performed on the to-be-processed information. The level of processing should be thought of as a point on a continuum rather than as a discrete level. The continuum goes from shallow processing, which focuses on perceptual features such as how a word sounds, to deeper processing involving meaning. Second, Craik and Lockhart assumed that the deeper the level, the more durable the resulting memory. If you need to remember an item for a long time, it is better to use a deep level of processing, such as focusing on its meaning, than to use a shallower level, such as how the item sounds. Third, the levels of processing view assumes that rehearsal can be relatively unimportant. Memory improvements are due to deeper levels of analysis, not to repeating an item over and over. Rehearsal will be beneficial only to the extent that it induces a deeper level of processing. This type of rehearsal, known as Type II or *elaborative rehearsal*, will improve memory; simple rote repetition, known as Type I or *maintenance rehearsal*, will not improve memory.

The final assumption has to do with how memory should be studied rather than with specifics of the theory. Because the emphasis is on processing rather than on structure, Craik and Lockhart (1972) argued that research will be most informative when the experimenter has control over the processing. When subjects are intentionally trying to learn, they may use the processing they consider most appropriate, which may not be the processing the experimenter wants. Therefore, researchers should use an *incidental learning* procedure. In this procedure, the subject is unaware that the material being processed will be tested later on; rather, the subject simply processes the material in the way suggested by the experimenter.

> The interesting thing to do is to systematically study retention following different orienting tasks within the incidental condition, rather than to compare incidental with intentional learning. Under incidental conditions, the experimenter has a control over the processing the subject applies to the material that he does not have when the subject is merely instructed to learn and uses an unknown coding strategy. (Craik & Lockhart, 1972, p. 677)

One implication is that intent to learn is not necessarily an important variable in memory research. Indeed, Postman (1964, p. 190) had earlier concluded the same thing, that "intent per se is not a significant variable in learning."

One early experiment that seemed to support many of the assumptions of the levels of processing view, including the relative unimportance of intentionality, was reported by Hyde and Jenkins (1973). They used 22 different conditions, but we will examine only a few in detail. The to-be-remembered items were 24 common words presented auditorily. Following presentation, the subjects were asked to recall as many of these items as they could in any order they liked (free recall). The groups of subjects differed primarily in the orienting instructions given to them and in whether they were told in advance of the free recall test.

One group of subjects was given a shallow task, such as checking the words for the presence of the letter *E* or *G*. It is easy to perform this task without having to process the meaning of the words. A second group was given the slightly deeper task of identifying the part of speech, such as noun or verb. A third group was asked to rate the frequency, or how common the word was. A fourth group was asked to rate each word on how pleasant or unpleasant it was. This is thought to be a particularly deep level of processing because not only do subjects have to process the meaning of the item, but they then have to consider the meaning in relation to whether it has pleasant associations. Half the subjects in each of these groups were told that they would also need to remember the items for a later test (the intentional group); the other half were told nothing about the memory test (the incidental group). Finally, a control group was told simply to remember the words for a later free recall test. The results are shown in Figure 5.1.

There are three particularly important findings. First, looking only at the incidental groups, the deeper the level of processing, the more words were recalled. For example, those subjects performing pleasantness ratings recalled more items than did those doing

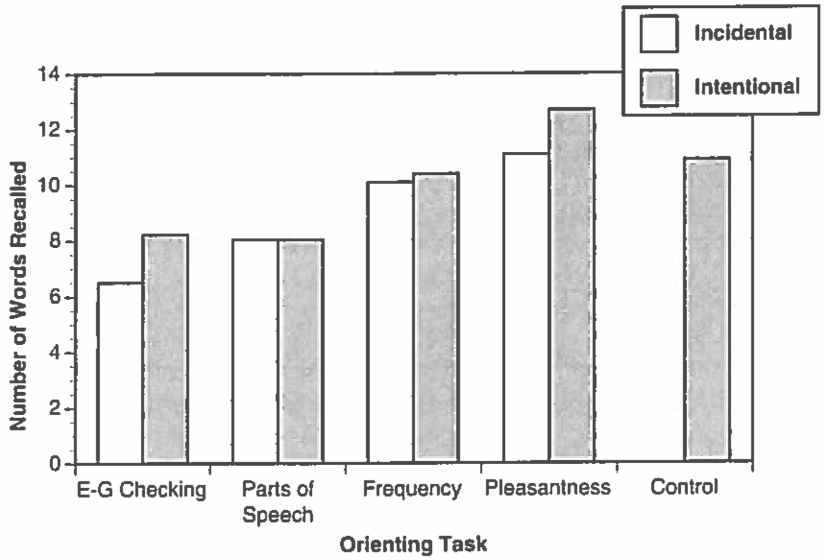

Figure 5.1 Amount recalled as a function of orienting task and whether subjects were told a memory test would be given. SOURCE: From "Recall of Words as a Function of Semantic, Graphic, and Syntactic Orienting Tasks," by T. S. Hyde and J. J. Jenkins, 1973, *Journal of Verbal Learning and Verbal Behavior, 12*, 471–480. Copyright © 1973 Academic Press, Inc. Reprinted with permission.

E-G checking. Second, subjects in the incidental pleasantness condition recalled as many, if not more, words than did subjects in the intentional control condition. This finding supports Postman's assertion that other factors may be more important than intent to remember. Finally, there was very little difference in performance between the incidental and intentional groups that were given the same orienting instructions. For example, those subjects performing the part of speech judgment recalled the same number of words regardless of whether they knew about the recall test.

The results from Hyde and Jenkins' (1973) study provide empirical support for many of the assumptions of Craik and Lockhart's (1972) levels of processing framework. A different set of experiments supported the idea that maintenance rehearsal has little or no effect on memory. Craik and Watkins (1973) tested the idea that maintenance rehearsal does not lead to improved memory. They induced subjects to use maintenance rehearsal by giving them special instructions. Subjects were asked to listen to a list of items and report the last word that began with a particular letter. If the critical letter was G, for example, the subject could ignore all words until one was encountered that began with G. Thus, in the list *daughter, oil, rifle, garden, grain, table, football, anchor, giraffe*, the first word the subject would rehearse would be *garden*. The subject would rehearse this item for only a moment before *grain* replaced it. *Grain* would be rehearsed for longer, because *giraffe* does not appear for a while. The operational definition of how long an item was rehearsed was termed *i*, the number of intervening words. *Garden* had an *i* value of 0, *grain* had an *i* value of 3, and some items had *i* values as large as 12. At the end of each trial, the subject would report the one item required, and the next trial would begin. The final manipulation of interest involved a surprise final recall test. Ten minutes after the final trial, Craik and Watkins asked the

Experiment *Levels of Processing*

Purpose: To demonstrate the basic levels of processing phenomena.

Subjects: Forty subjects are recommended. Ten should be assigned to the intentional E-G group, ten to the incidental E-G group, ten to the intentional pleasantness group, and ten to the incidental pleasantness group.

Materials: Table E in the Appendix lists 25 common two-syllable nouns. For each subject, present these in a different random order. The orienting responses require answer sheets with numbers from 1 to 25. Additional sheets will be required for the free recall test.

Design: Because each subject experiences only one condition, both intentionality and level of processing are between-subjects factors. The basic design is a 2 × 2 factorial.

Procedure: Inform the subjects that this is an experiment on information processing. Describe the orienting task that each group will get, and explain how they will respond on the answer sheets. Do not give out the response sheets for the free recall task; it is very important that the subjects in the incidental conditions do not expect a memory test. For those subjects in the intentional conditions, inform them that they will be tested for their memory for the words later on. Read each word aloud, allowing 3 seconds for the subject's response. After reading the words, remove the first answer sheet before giving them the second. This way, they won't be able to use their previous responses as cues. Allow as much time as needed for the free recall test.

Instructions for E-G: "This is an experiment on information processing. I will read a list of 25 words, and I would like you to indicate whether there is an E or a G in the word. If the word has either an E, a G, or both, please write *Y* next to the number on the response sheet. If the word has no Es or Gs, please write *N*. Please be as accurate as you can, but you will have only 3 seconds per word to make your response." To the intentional group, add, "At the end, you will be tested for your memory for the words."

Instructions for Pleasantness: "This is an experiment on information processing. I will read a list of 25 words, and I would like you to indicate whether you think the word represents something pleasant or unpleasant. If you think the word is pleasant, please write *P* next to the number on the response sheet. If the word is unpleasant, please write *U*. Please be as accurate as you can, but you will have only 3 seconds per word to make your response." To the intentional group, add, "At the end, you will be tested for your memory for the words."

Instructions for Test: "Please write down as many of the 25 words as you can remember. Feel free to write them down in any order."

Scoring and Analysis: For each subject, count the number of words recalled, and compute an average for each group. The results should be similar to those plotted in Figure 5.1. A 2 × 2 between-subjects analysis of variance can be performed on the individual subject data.

Optional Enhancements: Include a control group that receives intentional instructions but no orienting instructions.

SOURCE: Based on an experiment by Hyde & Jenkins (1973).

subjects to report as many of the critical items as they could—both the final ones, which they had already reported, and those that were not the final items. Craik and Watkins found that final recall was unrelated to the *i* value: Recall of an item was uncorrelated with how long it had been rehearsed using maintenance rehearsal.

This result poses a problem for the modal model because it suggests that time spent in short-term store can be unrelated to subsequent recall. Further research has generally replicated the finding of little or no effect of maintenance rehearsal. For example, Nairne (1983) replicated these results for free recall but also tested recognition. In a recognition test,

Table 5.1 Two ways of accounting for basic memory phenomena

Store	*Sensory*	*STS*	*LTS*
Format	Literal copy	Phonological	Semantic
Rehearsal	Not possible	Maintenance	Elaborative
Capacity	Medium	Small	Infinite
Duration	500–2000 ms	Up to 30 s	Minutes to years

NOTE: Reading the table down shows the structural view: If an item is in STS, then its format will be phonological, the capacity small, and so on. Omitting the top row shows the processing view: If an item is encoded phonologically, then the capacity is small and the duration is around 30 s. The first row becomes irrelevant.

subjects can indicate whether they do or do not recognize an item as being from a particular list. Nairne found increased "yes" responding with increased rehearsal, but this does not necessarily mean that overall memory was more accurate. Subjects were more likely to say "yes" both to items that were from the list (a correct response, known as a *hit*) and to items that were not from the list (an incorrect response known as a *false alarm*). When this tendency is taken into account, using signal detection theory and measuring accuracy with d' (see Chapter 9), there was no relation between rehearsal and performance.

Table 5.1 illustrates how the same results can be interpreted by two very different views. According to the modal model, it is the memory structure that determines the properties. If short-term store is used, the code will be phonological, the capacity will be relatively small, maintenance rehearsal will keep the item in the store, and without rehearsal, the maximum duration is about 30 s. What having levels of processing does, in effect, is eliminate the top row. According to levels of processing, if a subject processes an item at the phonological level, then the subject will not be able to remember very many items for very long.

Although levels of processing as a theoretical description can account for many experimental findings—including some findings that the modal model cannot—it has two major problems. The first major problem with levels of processing is that one of its key assumptions is *circular*. It is assumed that deeper levels of processing lead to better memory, and when there is better memory, we attribute it to a deeper level of processing. Although most researchers will concede that E-G checking is intuitively shallower than pleasantness ratings, there is no independent method for determining, before a memory experiment, whether process A is deeper or shallower than process B. Craik and Tulving (1975) reported ten experiments that attempted to define levels objectively. For example, one idea was that deeper levels of processing require more time to perform than shallower levels do. Unfortunately, they observed that it is possible to construct a situation such that a deep task requires 0.83 s to complete and leads to 82% correct, and a shallow task takes 1.70 s to complete but leads to only 57% correct.

The lack of an objective definition of a level might have been overcome had there not been a second major problem, this one concerning an omission. Levels of processing focuses almost exclusively on encoding and says relatively little about retrieval. The second major processing view developed in the 1970s, transfer appropriate processing, began as a way of rectifying this omission. Despite its major problems, however, levels of processing has had an enormous impact on the field and influenced many current theories of memory.

Transfer Appropriate Processing

The principal difference between levels of processing and transfer appropriate processing is that the latter includes retrieval. According to this view, a process leads to better performance not because it is deeper but because it is appropriate for the kind of test that will be conducted. In general, deep tasks do lead to better performance than do shallow tasks on free recall tests, but free recall is not the only way of testing memory. Morris, Bransford, and Franks (1977) demonstrated that a shallow level of processing can lead to better performance than a deep level when tested in a particular way. This demonstration clearly disconfirms an assumption of levels of processing.

Morris, Bransford, and Franks (1977) presented 32 sentences auditorily to the subjects, and in each sentence, a word was missing. In one condition, subjects judged whether a target word filled in the blank. For example, the sentence might be "The (blank) had a silver engine" and the target word might be *train*. In a different condition, subjects judged whether the target word rhymed with another. In this case, the sentence might be "(Blank) rhymes with legal" and the target word might be *eagle*. Subjects indicated "yes" or "no" to each kind of question and were unaware that a later memory test would be given. So far, this is just like a standard levels of processing experiment, with subjects performing incidental learning under two different sets of orienting instructions. But the researchers changed the paradigm by including two types of tests. One test was a standard recognition test, in which a target word was presented and subjects were asked whether it had been seen previously. The second test was a rhyming recognition test, where subjects were asked whether a word rhymed with one of the target words. In the example given above, the standard recognition test might have as a test item the word *train* or the word *eagle*. The rhyming recognition test might have as test items the word *plane* (which rhymes with *train*) and *regal* (which rhymes with *eagle*). The results are shown in Figure 5.2.

With a standard recognition test, the deeper orienting task (whether the word fit in the sentence) led to better performance than did the shallower orienting task (whether the word rhymed). However, with the rhyme recognition test, the shallow task led to better performance than did the deep task. Levels of processing assumed that deeper processing leads to better performance, but the data in Figure 5.2 show that this is true only for some kinds of tests; on other tests, shallow processing can be better than deep. Morris, Bransford, and Franks (1977) interpreted their results as supporting a theory called *transfer appropriate processing*. A type of processing will lead to better memory performance if it is appropriate for the particular test; no one type of processing is good for all tests.

One qualification should be added. Fisher and Craik (1977) replicated the results of Morris, Bransford, and Franks (1977), ruling out several uninteresting interpretations of the data. They emphasized, however, that although rhyme encoding followed by a rhyme cue led to better performance than did semantic encoding followed by a rhyme cue, semantic encoding followed by a semantic cue led to the best performance. This can be seen in the data shown in Figure 5.2. According to Fisher and Craik (1977), the reason is that semantic encoding leads to a more specific cue, which in turn leads to enhanced discriminability from competing cues. For example, there is only a limited number of possible phonemic cues, compared to an almost infinite number of semantic cues. This is related to Watkins' (1979) *cue overload principle*, which states that only so many items

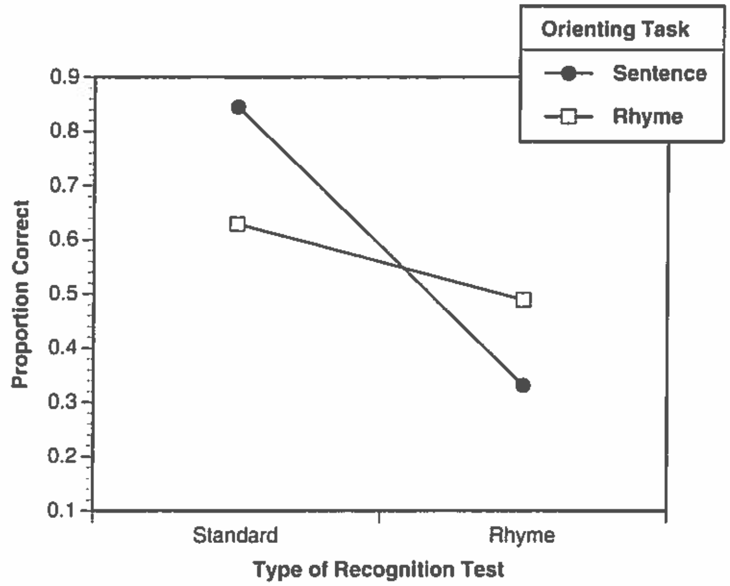

Figure 5.2 Overall proportion correct on a standard or rhyme recognition memory test following either a sentence (semantic) or rhyme (nonsemantic) orienting task. SOURCE: From "Levels of Processing Versus Transfer Appropriate Processing," by C. D. Morris, J. D. Bransford, and J. J. Franks, 1977, *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533. Copyright © 1977 Academic Press, Inc. Reprinted with permission.

can be associated with a cue before the cue begins to lose its effectiveness. The more items that are associated with a cue, the less effective the cue will be in eliciting the desired item.

Organization and Distinctiveness

The levels of processing framework includes the idea that a deeper level of processing produces a more distinctive cue than does a shallow level of processing. The more distinctive an item, the more different it is from competing items (see also the discussion of distinctiveness in Chapter 6). At the same time, however, a large body of literature demonstrates that organization helps memory. When incorporating information into an organization, what matters most are the similarities. As Hunt and McDaniel (1993, p. 421) wondered, "How can both similarity and difference be beneficial to memory?"

As used in the memory field, *organization* refers to relationships among the information that is to be remembered, and it formed a cornerstone of Gestalt theory (Katona, 1940; Koffka, 1935). Subjects recall more words from a categorized list than from an uncategorized list. For example, Deese (1959a) found that subjects recalled about 49% of the words from a list that contained high associates (such as *flower, insect, bees*), compared to only 37% of the words from a random list. The effects of organization can be quite powerful. Mandler (1967) had four groups of subjects examine 52 words. Of most interest here is that subjects who were asked to sort the cards into categories of their own choosing recalled as many of the words on a surprise test as did subjects who were specifically asked to memorize the words.

A related phenomenon is called *clustering*. Bousfield (1953) presented subjects with a 60-word list made up of 15 instances of four categories (animals, names, professions, and vegetables). The words were presented in random order. Bousfield observed that subjects were likely to cluster their responses; that is, they were likely to recall two or more items from the same category sequentially even though the items were separated in the list. The lists do not need to be categorized in such a blatant fashion in order to see clustering (for example, Tulving, 1962).

One resolution has been to say the paradox that both similarity and difference helps is more apparent than real. In other words, both are important because both reflect particular types of processing. Hunt and his colleagues (Hunt & Einstein, 1981; Hunt & McDaniel, 1993) emphasize the importance of relational and item-specific processing, arguing that organization emphasizes relational processing (how the items fit together), whereas item-specific processing emphasizes the particular to-be-remembered item. Memory performance is best when both processes occur. A specific example is provided in Chapter 13 at the end of the section "Item Versus Order Information."

For now, the important point is that both relational and item-specific processing are important. Without knowledge of the to-be-remembered item itself, memory will be quite poor. Even if the item is known, however, relational processing can help provide cues that quickly produce the desired information. This interaction between the processes performed at encoding and those performed at retrieval is highlighted in our next topic, the encoding specificity principle.

The Encoding Specificity Principle

According to the *encoding specificity principle* (Tulving, 1972, 1983), the recollection of an event or a certain aspect of it depends on the interaction between the properties of the encoded event and the properties of the retrieval information. This idea is related to transfer appropriate processing in that both emphasize the interaction between the processing that occurred during encoding and the processing that occurs at retrieval. Tulving refers to the interaction as *ecphory*, from a Greek word meaning "to be made known." An ecphoric process, then, is a process by which "retrieval information is brought into interaction with stored information" (p. 178). Because it is the interaction that is important, no statements can be made about the mnemonic properties of an item, a process, or a cue (p. 239). For example, one cannot say that deeper processing leads to better recall than shallow processing because recall depends on the type of processing at retrieval. It is only when the interaction of the encoding and retrieval conditions is specified that meaningful statements can be made.

CogLab Experiment
http://coglab.wadsworth.com/experiments/EncodingSpecificity/

We have already examined one good example: The statement "deep processing at study is better than shallow processing" is inaccurate unless the statement is amended to

Table 5.2 The probability of recall when study cues and retrieval cues are varied

	Retrieval Cue		
Study Cue	*None*	*Weak*	*Strong*
None	0.49	0.43	0.68
Weak	0.30	0.82	0.23

SOURCE: Data from Thomson & Tulving (1970).

include a description of the test. Another example concerns the difference between a strong cue and a weak cue. A *strong cue* is a word that elicits a target word most of the time. When people hear the word *bloom*, most will respond with *flower* as the first word they think of. A *weak cue* is a word that elicits the target word only rarely. When people hear *fruit*, they respond with *flower* only about 1% of the time. You might be tempted to conclude that strong cues are better than weak cues, but this is the sort of absolute statement that Tulving (1983) argues is meaningless. The reason is that weak cues can be made more effective than strong cues depending on what happens at study and at test. In one experiment, Thomson and Tulving (1970) presented some target words alone at study, and some with a weak cue. At test, there were three cue conditions: no cue, a weak cue, or a strong cue. The results, expressed as probability of recall, are shown in Table 5.2.

In the control condition, in which there were no study cues and no retrieval cues at test, the probability of recalling the target word was 0.49. When there was no cue present at study, a strong cue at test led to better performance (0.68) than either a weak cue (0.43) or no cue (0.49). However, when there was a weak cue present at study, the weak cue at test led to better performance (0.82) than a strong cue (0.23). The important result here is that the effectiveness of even a long-standing strong cue depends crucially on the processes that occurred at study.

These results have two important implications. First, and perhaps most surprising, is that an item is not necessarily the best cue for itself. As we shall see in more detail in Chapter 9, there is a phenomenon known as *recognition failure of recallable words* in which the subject is asked to recall items that were on a list and also to recognize which items were on the list. In this paradigm (e.g., Watkins & Tulving, 1975), the subject studies a list of word pairs in which the first word in the pair is a weak cue of the second word; for example, *glue* might be paired with *chair*. Then a word that is a strong cue is given, such as *table*, and subjects are asked to write down as many words as they can that are associated with *table*. A recognition test is then given, in which a copy cue (the word *chair* itself) is used, and subjects often fail to recognize this item as one of the words on the original list. The final test is cued recall, where *glue* is the cue. Now, subjects are quite likely to recall the word they earlier failed to recognize. *Chair* is not always an effective cue when trying to remember *chair*. This finding can be explained only by theories that take into account processing at both encoding and retrieval.

The second implication is perhaps less surprising, but involves transfer appropriate processing. As Nairne (2001) has pointed out, many people have taken transfer appropriate processing to mean that performance is better when the encoding and retrieval conditions match than when they mismatch. It is true that in many cases, such as with the context-dependent memory effects discussed next, it is the match between encoding and retrieval that seems important. However, there are two things wrong with focusing on the similarity or match between the processes performed at encoding and retrieval. First, such a description is not transfer appropriate processing, but a different theory, transfer similar processing. All that transfer appropriate processing requires is that the processes be appropriate, not that they be similar. They can, of course, be similar, but that is not necessary. Second, transfer similar processing is empirically incorrect. In the Thomson and Tulving (1970) study, it is true that performance was best when there was a weak cue at study and a weak cue at test. However, there is another "match" condition: when no cues are presented at study and no cues are presented at test. Performance in this second "match" condition was worse than in the "mismatch" in which there are no cues at study and a strong cue at test. The results of Thomson and Tulving (1970) demonstrate that it is not the match that is important, but the appropriateness of the processing.

One reason many people have suggested that the match between encoding and retrieval is important is that many times similar processing is appropriate. We now examine this idea in greater depth by focusing on one paradigm.

Context and Memory

Many early experimental psychologists studied context effects or, more accurately, the change in performance when information is studied in one context and remembered in a different context (e.g., Carr, 1925; McGeoch, 1932). For example, Abernathy (1940) demonstrated that students who were tested in the same classroom in which they were taught did better on tests than did students who were tested in a different classroom. Such effects are predicted by the encoding specificity principle and, more generally, by any view of memory that emphasizes processing.

You should not be unduly worried about this result; changing classrooms from study to test often produces no decrement (e.g., Saufley, Otaka, & Bavaresco, 1985) because of at least three factors. First, most classrooms are quite similar; thus, a change from one to another will not be a very dramatic change. Second, only a limited amount of studying occurs in the first classroom, with much taking place in libraries, dorm rooms, and other locations. Third, students are taking the tests for a grade in a course, and so it should not be surprising that they do everything they can to do well. Even if you are worried, there are ways to overcome potential context effects.

Smith (1979) had subjects learn 80 words in a distinctive basement room. They were then given a recognition test on some of the words so that they would think that this phase of the experiment was over and would not rehearse the words during the next 24 hours. The next day, all subjects were given a surprise free recall test. One group was tested in the same basement room (same context), one was tested in a soundproof booth located in a computer room on the fifth floor (different context), and the third group was also tested in the booth but was given additional instructions. This third group of subjects was told to try to reinstate their memory for the basement room before trying to recall any items; they were asked to list attributes of the original room, and then try to recall the words. The same-context group recalled 18 words, on average. The different-context group recalled significantly fewer items, only 12. The third group, the different-context-plus-image group, recalled 17.2 items—almost the same number as the first group. It appears that environmental contextual information can be reinstated and can be used to compensate for a change in location.

Context has been defined in many different ways, although all share a similar distinction. For example, Wickens (1987) distinguishes between context alpha and context beta: *Context alpha* refers to "the environmental surrounds in which some event exists or occurs. . . . [There is] no implication that the context or the environment influences the event or is related to it in any significant way." *Context beta* is the "situation in which one stimulus event combines with another stimulus event to define the correct response or meaning of the event" (pp. 135–146).

The lack of an effect for changing classrooms notwithstanding, *environmental context-dependent memory* effects, in which context alpha is manipulated, are pervasive (Smith & Vela, 2001). Perhaps the most famous experiment that manipulated context alpha was reported by Godden and Baddeley (1975). Their subjects were members of a university diving club who were asked to learn a list of words either on land or 20 feet under water. They were then tested either on land or under water with four conditions: two match conditions, in which the environment (either on land or under water) was the same at both learning and testing; and two mismatch conditions, in which the environment was different at learning and testing. Regardless of where they were learning or recalling, the subjects wore their diving apparatus. There was a 4-minute interval between study and test, to allow enough time for the subjects to change environment. The results are reproduced in Figure 5.3. Subjects clearly recalled more words when the study and test environments matched than when they mismatched, regardless of whether subjects were under water or on land. When there was a mismatch, recall suffered. One practical impli-

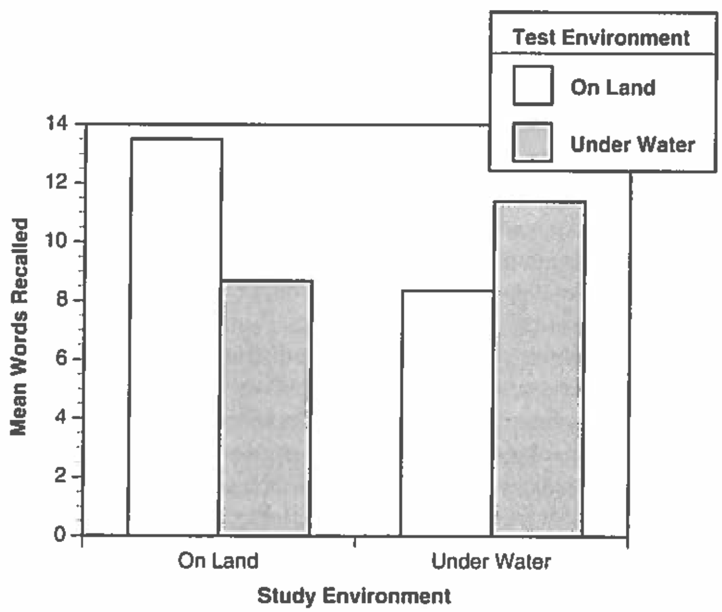

Figure 5.3 Mean number of words recalled as a function of study and test environments. SOURCE: From "Context-Dependent Memory in Two Natural Environments," by D. R. Godden and A. D. Baddeley, 1975, *British Journal of Psychology*, 66, 325–331. Copyright © 1975 British Psychological Society. Reprinted with permission.

cation for divers is that if they memorize dive tables on land, they may not remember them very well when needed (e.g., Martin & Aggelton, 1993). A better instructional technique is to have the divers learn the material under water as well as on dry land.

One potential confound might be that the actual movement necessary to change environments was causing the memory deficit, rather than the changed environment itself. To rule out this possibility, Godden and Baddeley conducted a second experiment. Half of the subjects again learned a list of words and were then tested for recall on land. The other half also learned the list and were tested on land, but they were required to enter the pool, swim a short distance, dive to a depth of 20 feet, and then return to land prior to recall. If the movement itself was the cause of the deficit, then there should be a difference in performance between the two groups. If the change in environment caused the deficit, then there should be no difference. There was no difference; therefore, the disruption between study and test did not cause the decreased recall.

An example of context beta comes from a study by Light and Carter-Sobell (1970). They presented to subjects sentences that contained a two-word phrase that was capitalized and underlined, such as "The <u>STRAWBERRY JAM</u> tasted great." Subjects were told to expect a memory test for the underlined and capitalized phrases. After all the sentences had been presented, a recognition test was given. Half of the phrases were old, meaning they were ones the subjects had seen, and half were new, meaning the subjects had not seen them. Some of the old phrases contained the same noun as the original sentence, but now the noun was paired with a new adjective, such as <u>TRAFFIC JAM</u>. The subjects were supposed to respond on the basis of the noun. Thus, if they saw <u>JAM</u>, they were to indicate that it was one of the studied items, regardless of whether the adjective was the same or different. When <u>JAM</u> was presented in the same context, subjects correctly recognized it about 65% of the time, but when it was in a new context, they recognized it only about 25% of the time. This is an example of context beta because the adjective was essential for determining the meaning of the word.

A *context-dependent memory effect*, then, refers to the observation that memory can be worse when tested in a new or different context relative to performance in a condition in which the context remains the same. Context can be manipulated in a variety of ways. For example, Schab (1990) filled a small room with the odor of chocolate either at study, at test, or at both. Even after a 24-hour delay, subjects' performance on a surprise test was better when the odors matched. This experiment also provides empirical validation for the famous anecdote reported by Marcel Proust in the first novel of *Remembrance of Things Past*.

State-dependent memory refers to a finding similar to context-dependent memory; indeed, many theorists view them as essentially the same (e.g., Capaldi & Neath, 1995). A person's state can be changed by altering the affective or the pharmacological state from study to test. The first person to describe this concept—that what has been learned in a certain state of mind is best remembered in that state—was a French aristocrat, Marquis de Puységur, in 1809 (Eich, 1989). More recently, Morton Prince (1910) conjectured that people have difficulty remembering their dreams when awake not because they do not want to, but because of the dissimilarity of the two states. In 1934, Susukita demonstrated that something learned while calm cannot be recalled when in a state of shock, but can be recalled at a later time when calm is restored. In this particular instance, a state of shock was induced when an earthquake hit. Although the typical finding is similar, some researchers vary the environment, some the pharmacological state, and some the affective state or mood of the person.

Table 5.3 Mean number of errors of recall as a function of encoding and testing state

	Retrieval	
Encoding	*Sober*	*Intoxicated*
Sober	1.25	2.25
Intoxicated	4.58	2.50

SOURCE: Goodwin, Powell, Bremer, Hoine, & Stern (1969).

Bartlett and Santrock (1979) altered the affective state of their subjects so that they were in either a happy or a neutral mood at study, and a happy or a neutral mood at test. Performance was better when the moods matched than when they mismatched. Similar results have been reported by Bower (1981). Although there are several well-documented failures to find mood-dependent memory effects, a review by Eich (1995) goes a long way toward delineating when such effects will or will not be seen. First among these conditions is producing a strong, sincere mood.

Goodwin, Powell, Bremer, Hoine, and Stern (1969) demonstrated state-dependent memory using alcohol. Some subjects received 10 oz of 80-proof vodka either at study, at test, or at both; other subjects remained sober. Not surprisingly, performance was best when subjects were sober at both study and test. Perhaps more surprising is the finding that the worst performance was when subjects were intoxicated at study and sober during the test. In fact, subjects were worse in this condition than when the mismatch was the other way around—sober at study and intoxicated at test (see Table 5.3). Similar state-dependent memory effects have been shown using marijuana (Eich, Weingartner, Stillman, & Gillin, 1975) and nicotine (Peters & McGee, 1982).

Weingartner and Faillace (1971) conducted a state-dependent memory experiment with two different groups of subjects. One group was made up of chronic alcoholics with documented histories of long-term alcohol abuse; the other group was made up of an equal number of nonalcoholics who were closely matched to the alcoholic subjects with respect to age, education, and other demographic variables. The test session was held two days after initial learning, during which subjects tried to recall as many of the target words as they could. The subjects from both groups were either intoxicated or sober at study, and intoxicated or sober at test. In this experiment, intoxication was defined as having 1.6 ml of alcohol per kg of body weight. Both groups of subjects showed state-dependent memory effects.

State-dependent memory shows an interesting asymmetry in which a shift from intoxication to sobriety impairs memory more than a shift from sobriety to intoxication (Eich, 1989; Overton, 1984). For example, as Table 5.3 shows, there is a larger difference between recalling items encoded during intoxication while in an intoxicated state than while in a sober state (mean errors of 2.50 versus 4.58), and a smaller effect of recalling items encoded during sobriety while in a sober state than while in an intoxicated state (mean errors of 1.25 versus 2.25). As Eich (1989) points out, this asymmetry can be seen in the studies of Eich et al. (1975) with marijuana, Peters and McGee (1982) with nicotine, and Bartlett and Santrock (1979) with mood. Furthermore, it is seen in the Weingartner and Faillace (1971) study in both the alcoholic and the nonalcoholic control groups. If we label the sober state as normal for the nonalcoholics and the intoxicated state as normal for the alcoholics, then the pattern is identical: Performance is worse when switching from the unusual to normal condition than when switching from the normal to unusual condition, regardless of which condition (sober or intoxicated) is normal.

The interpretation advanced by several theorists to explain this asymmetry is actually quite simple (e.g., Barry, 1978; Capaldi & Neath, 1995; Eich, 1980). The context in which a person—or other animal, because all these context- (alpha and beta) and state-dependent memory effects can be observed in dogs, rats, and pigeons (e.g., Asratian, 1965; Bouton, 1993; Rescorla, Durlach, & Grau, 1985)—learns some information provides cues that the subject can use to remember the information (see Chapter 6 for more detail). Some N (or normal) cues will always be encoded because there will always be some aspect that is unchanged. It can be assumed that drugs add U (or unusual) cues to the representation, with N or normal state cues being little affected. For chronic alcoholics, being sober is unusual, and so the absence of the effects of alcohol provides the U cues. The N-to-N and U-to-U groups show generally good retention because the cues encoded at study (either N or U) are the cues that are useful at test. The N-to-U group shows an intermediate level of performance because some N cues were present at test as well study. The U-to-N group shows the worst performance because the U cues present at study are absent at test.

One advantage of this view is that it makes a specific prediction: In an experiment in which context is changed, the change in context can be overcome if the experimenter provides useful cues to the subject. We have already seen one example, in which the experimenter told the subject to reinstate the missing context mentally (Smith, 1979). This can be thought of as providing some of the missing U cues. Another example comes from a study reported by Eich and Birnbaum (1982), in which subjects were asked to learn a list of words in either a sober or an intoxicated state. When tested after a change in drug state, retrieval was poor, as is usually the case. However, some of the subjects in the changed-state condition were given additional cues (for example, "Think of all the flowers that you can"). Providing a category name gave the subject additional useful cues, and recall was much improved (Eich & Birnbaum, 1982).

Note that in almost all of these cases, a recall test rather than a recognition test was used. There are many reports of failures to obtain a context-dependent memory effect when recognition is used (e.g., Godden & Baddeley, 1980; Smith, Glenberg, & Bjork, 1978), as well as reports of failures to obtain mood-dependent memory effects with recognition (Eich & Metcalfe, 1989). Smith and Vela (2001) conducted an analysis of 75 experiments from 41 articles published between 1935 and 1997 and concluded that environmental context effects are reliably found, and can also be seen in recognition tests. To the extent that the processing at study and at test emphasizes noncontextual cues, or that reinstatement of the original context is encouraged, environmental context effects will be reduced.

Context-, mood-, or state-dependent memory effects, then, are larger (1) with tests such as free recall that require internally generated cues than with tests like cued recall or recognition that also have externally provided cues; and (2) when the tests emphasize conceptually driven tasks (those that focus on word meanings and relationships between words) rather than data-driven tasks (those that stress perceptual processing). This latter aspect is nicely illustrated in a series of experiments reported by Eich and Metcalfe (1989). Before describing these experiments, however, it is necessary to describe some methods and define some terms.

Johnson and Raye (1981) defined *cotemporal thought* as "the sort of elaborative and associative processes that augment, bridge, or embellish ongoing perceptual experience but that are not necessarily part of the veridical representation of perceptual experience" (p. 443). Eich and Metcalfe suggested that if a person is in a happy mood, then the cotemporal thought that occurs should lead to large mood-dependent memory effects if the cotemporal thought is allowed to be important. To test this hypothesis, they used a procedure, described first by Heller (1956) and more recently by Slamecka and Graf (1978), that gives rise to the *generation effect* (see Chapter 13 for more discussion).

In the typical generation-effect experiment, on half of the lists subjects are asked to read an item, and on the remaining trials they are asked to generate an item. For example, subjects may see a category cue followed by two words, one of which is in uppercase. Subjects are instructed that they will need to remember the uppercase word for a later memory test. In the generate condition, subjects see exactly the same display except that only the initial letter is given for the uppercase word: The subjects have to generate the word themselves. The typical finding is that generated words are remembered much better than are the read words.

Milkshake flavors: chocolate : VANILLA	[Read condition]
Milkshake flavors: chocolate : V———	[Generate condition]

Eich and Metcalfe (1989) combined a generation-effect procedure with a mood-manipulation procedure. Half of the subjects were induced to be happy at study, and half were induced to be sad. Of the happy subjects, half were induced to be happy again 24 hours later at the surprise recall test, and half were induced to be sad. Within the four groups delineated by mood, half of the subjects read the target item, and half generated the target item. Mood changes were induced by playing either happy (Mozart's *Eine Kleine Nachtmusik*) or sad (Albinoni's *Adagio in G minor*) music and instructing subjects to think either happy or sad thoughts. The results are displayed in Figure 5.4.

Subjects recalled more words when their moods matched than when the moods were mismatched, regardless of whether they read or generated the items. However, when the words were generated, not only was recall superior, but the size of the mood congruency effect was larger. These data, and others reported in the same article, support the idea that, to the extent that thoughts are important in the processing of information, there will be large effects of environment, state, and mood congruency. If thoughts are relatively unimportant, then context effects will be difficult to obtain.

Context- and state-dependent memory can be seen as examples of transfer appropriate processing. When the processing done at study (such as relying on U cues) is not appropriate for the processing required at test (generating a response based primarily on N cues), memory will be impaired. The addition of the missing cues, however, can dramatically increase the amount of information remembered. This type of finding—increased recall under one set of testing conditions relative to another set of testing conditions—has important implications for forgetting, a subject we turn to in the next chapter.

There is an important distinction between mood-dependent memory, which is shown in Figure 5.4, and mood-congruent effects. *Mood congruency* refers to the finding that a given mood tends to cue memories that are consistent with that mood, rather than memories that are inconsistent (Blaney, 1986; Hertel, 1992). For example, Teasdale and Russell (1983) had subjects learn a list of negative, neutral, or positive words. At test, subjects were induced to be either happy or sad and then were asked to recall as many words

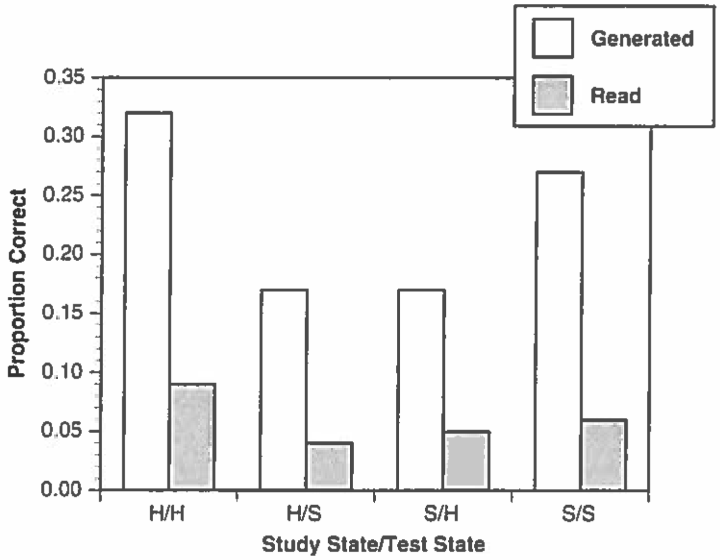

Figure 5.4 Probability of correct recall as a function of mood at study and test (happy or sad) and whether the item was generated or read.
SOURCE: From "Mood-Dependent Memory for Internal Versus External Events," by E. Eich and J. Metcalfe, 1989, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 443–455. Copyright © 1989 American Psychological Association. Reprinted with permission of the author.

as they could. The sad subjects recalled more negative words than the happy subjects did, but the happy subjects recalled more positive words than the sad subjects did. The reason for mood-congruency effects is that the mood serves as a cue: Positive events are associated with positive moods, and negative events with negative moods. A particular type of mood can cue memories that are associated with that mood.

The Process Dissociation Framework

The levels of processing and transfer appropriate processing views stimulated a lot of research on the effects of different kinds of processing. One potential problem concerns how to attribute specific effects to specific processes. Several processes are used at the same time, and until quite recently it appeared difficult to separate out the effects of different operations. Jacoby (1991) has developed a technique for separating different processes that may be operating. So far, it has been applied to only a few types of processes, but it should be relatively straightforward to apply the same logic to other settings.

Jacoby and his colleagues (Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993) were interested in separating intentional retrieval processes from incidental processes. (This topic is of critical importance in the implicit memory literature and will be discussed further in Chapter 7.) On many tasks, performance improves simply because an item has been perceived a little earlier. Even though the subject may be completely unaware of this influence, performance can be enhanced. For example, if you happened to read the word *toboggan*, and a little later were doing a crossword puzzle in which *toboggan* was an answer, you would be more likely to get the answer than someone who had not recently read the word. This sort of process is sometimes termed an *automatic* or *incidental process*. By contrast, an *intentional process* is one in which the subject intentionally tries to recall an item. Because both of these processes usually contribute to test performance (Richardson-Klavehn & Bjork, 1988), it has been difficult to determine the properties and effects of each.

One way to separate them, Jacoby (1991) suggests, is to have two groups of subjects study a set of words. One group receives an inclusion test, in which both processes are allowed to contribute; the other group receives an exclusion test, in which only automatic or incidental processes can contribute. The retrieval cues available to the subject are kept constant; the sole difference between an inclusion test and an exclusion test is the instructions given to the subject. By comparing performance on the two tests, it becomes possible to separate the effects of each process.

To see how this logic works, it is easiest to follow a specific example. In Phase 1, all subjects hear a list of items. In Phase 2, more items are presented, but this time in one of two conditions. In the full-attention condition, the subjects simply read the words. In the divided-attention condition, the subjects also listen to a tape-recorded series of numbers and indicate when they hear three odd digits in a row. The divided-attention manipulation is designed to reduce the influence of the recollective process while leaving the automatic process unaffected. At test, the subjects receive a stem completion task, in which they are asked to complete the stem to make a valid English word. The stems are presented in one of two colors, green or red. If the subjects see a green stem, they are asked to use it as a cue to remember a word from either Phase 1 or Phase 2; if they cannot remember a word, they are asked to complete it with the first word that comes to mind. A green stem, then, signals an inclusion test, on which both automatic and intentional processes work together. If the subjects see a red stem, they are asked to use it as a cue to remember a word from either Phase 1 or Phase 2 but to make sure that they do *not* write any of these words as a response. A red stem signals an exclusion test, on which automatic and intentional processes work in opposition.

Table 5.4 shows the probability of responding with an old word, from two versions of this experiment. Note that in the exclusion condition, subjects were instructed not to write down any old items. In general, performance is worse under divided than under full attention, particularly for the inclusion task. Performance is equivalent for the "heard"

Table 5.4 The probability of completing the stem with a studied item on both inclusion and exclusion tests for experiments

	Performance Component			
	Read		*Heard*	
Attention	*Inclusion*	*Exclusion*	*Inclusion*	*Exclusion*
Full	0.61	0.36	0.47	0.34
Divided	0.46	0.46	0.42	0.37

NOTE: "Read" and "Heard" correspond to whether the item came from Phase 1 or Phase 2 of presentation.

SOURCE: Jacoby, Toth, & Yonelinas (1993).

words because attention was not manipulated there. This is an important condition because it provides a check on the attention manipulation: If performance did not differ between read and heard words, we would not have confidence in the division of attention manipulation.

How can we separate the recollective process from the automatic process using these data? On an *inclusion* test, the probability of responding with a studied item (either read or heard) is the probability of conscious recollection (R) plus the probability that a word came to mind automatically (A) when there is a failure of conscious recollection ($1 - R$). Because the second part is a conditional probability, we need to multiply the two components together to reflect the fact that both need to occur.

$$\textit{Inclusion} = R + A(1 - R) \tag{5.1}$$

For an *exclusion* test, the word will be produced only when conscious recollection fails ($1 - R$) and when the word automatically came to mind (A).

$$\textit{Exclusion} = A(1 - R) \tag{5.2}$$

Because Equations 5.1 and 5.2 share similar terms, Equation 5.2 can be rewritten as

$$\textit{Inclusion} = R + \textit{Exclusion} \tag{5.3}$$

or

$$R = \textit{Inclusion} - \textit{Exclusion} \tag{5.4}$$

If we divide Equation 5.2 by ($1 - R$), we can solve for A:

$$A = \frac{\textit{Exclusion}}{1 - R} \tag{5.5}$$

Using the values in Table 5.4, we can now calculate estimates of R and A for the words that were read. We focus on this condition because it was only here that the divided-attention manipulation was performed. The estimate of R for the full-attention condition in this experiment is calculated using Equation 5.4: $R = 0.61 - 0.36$, or 0.25. The estimate of A for the same condition is calculated using Equation 5.5: $A = 0.36/(1 - 0.25)$, or 0.47. For the divided-attention condition, $R = 0.46 - 0.46$, or 0, and $A = 0.46/(1 - 0)$, or 0.46. These values are summarized in Table 5.5.

The estimates of R and A suggest that the processing dissociation technique can separate the influences of two different processes. Divided attention was designed to minimize the effect of conscious recollective processes, because the subject was required to attend

Table 5.5 Estimates of the contribution of recollective and automatic processes for the words that were read, calculated from the data in Table 5.4

	Estimate	
Attention	*Recollection*	*Automatic*
Full	0.25	0.47
Divided	0.00	0.46

SOURCE: Jacoby, Toth, & Yonelinas (1993).

to the irrelevant vigilance task. As Table 5.5 shows, the contribution of *R* was reduced to 0. However, attention should not affect the automatic component, and indeed the estimates of *A* are identical in both the divided- and full-attention conditions. Jacoby, Toth, and Yonelinas (1993) performed several more replications to ensure that the results obtained here were not a fluke. In particular, in a second experiment, they again obtained an estimate of *R* in a divided-attention condition of 0.0.

There have been some criticisms of this approach, centering around the issue of independence of the intentional and automatic processes. For example, Richardson-Klavehn, Lee, and Joubran (1994) suggest that the two processes may operate in a symbiotic fashion: An item suggested by the automatic process may then be considered by the intentional process. The question of independence of these processes is explored in detail in an exchange between Jacoby and his colleagues (Jacoby, Begg, & Toth, 1997; Jacoby & Shrout, 1997) and Curran and Hintzman (1997; Hintzman & Curran, 1997). It is unclear at the moment whether this criticism is important or trivial; at least one study (McBride, Dosher, & Gage, 2001) found equivalent results when using a version that assumed independence and one that did not. Regardless, the usefulness of the procedure is becoming more apparent. The procedure has been used in an increasing variety of tasks, including applications to implicit learning (Destrebecqz & Cleeremans, 2001; see Chapter 7), amnesia (Verfaellie & Treadwell, 1993; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998; see Chapter 8), and aging (Hay & Jacoby, 1999; Jennings & Jacoby, 1993; see Chapter 14). Further confidence comes from experiments in which the estimates obtained by process dissociation agree with estimates from other procedures (Yonelinas, 2001).

What Is Encoded?

The studies on state-, mood-, and context-dependent memory suggest that even when presented with a single word, people also encode many other attributes related to their mood and their environment. Furthermore, as we shall see in the next chapter, people also seem to encode all kinds of information about the word itself, including whether it is a noun or an adjective, whether it has one or two syllables, whether it was presented in red or in green, whether it was presented in upper- or lowercase, and so on. Is this much information really being encoded? One possible answer is that an enormous amount of processing, encoding information on a large number of dimensions, occurs even in the 500 ms that a word is available for study.

There is a different way of viewing this question, however. Many of the dimensions that subjects appear to encode may be apparent at test only when particular comparisons are made. For example, describe the object in Figure 5.5A. Most people will say "a square." However, when asked to describe the same object when placed next to another item, as in Figure 5.5B, they now describe it as a small square. With two other items, as in Figure 5.5C, they describe it as a small white square.

Thus, there is a difference between a nominal stimulus and a functional stimulus. The *nominal stimulus* is what the experimenter thinks the subject is encoding, and the *functional stimulus* is what the subject actually encoded. In general, the functional stimulus is likely to include many elements of the context and many idiosyncratic associations. Taking heed of an assumption from levels of processing, memory is a by-product of processing information, and the processing can add to or subtract from the nominal stimulus.

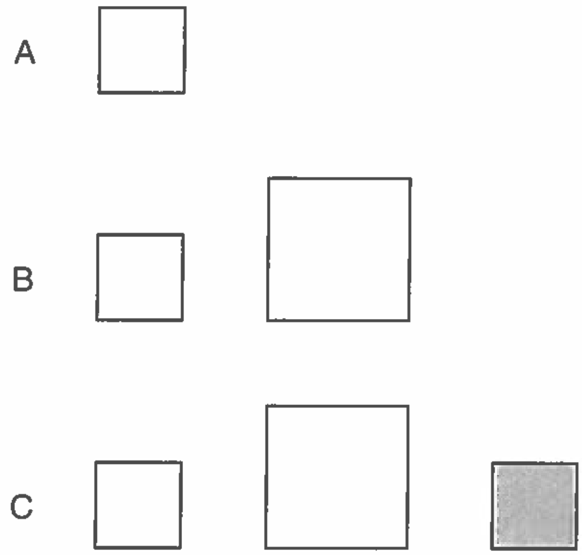

Figure 5.5 An illustration of how different dimensions become relevant. When asked to name the figure in A, the answer is a square; in B, the answer is a small square; and in C, the answer is a small white square.

One famous example of this process is the study reported by Carmichael, Hogan, and Walter (1932; see also Moore, 1910). The subjects were told that they would see a picture and would be asked to reproduce the figure as accurately as possible. As each figure was shown, the experimenter would say, "The next figure resembles . . ." and would say a name. For example, one group of subjects was told that an ambiguous figure (such as those shown in the center column of Figure 5.6) resembles "eyeglasses," and a second group was told that it resembles "dumbbells." Although the task was to reproduce the figure "as accurately as possible," most subjects changed the figure so that it more closely resembled the verbal label. Examples of some reproductions are also shown in Figure 5.6.

The point is, of course, that the nominal stimulus is the figure itself, but the functional stimulus—how the subject encoded the item—was slightly different. The subject paid attention to the verbal label, and this affected how the figure was encoded. Similar effects are observable when the labels are provided only at test (Hanawalt & Demarest, 1939). The way the information is processed at study or at test can distort the original information. (We shall examine reconstructive processes in more detail in Chapter 12.)

Chapter Summary

The levels-of-processing framework was the first detailed account of memory to argue that the type of processing was more important than the underlying theoretical structure. According to this view, items will be well remembered if they are processed deeply, regardless of factors such as intent to learn. The primary difference between levels of processing and transfer appropriate processing is that the latter specifically includes retrieval factors.

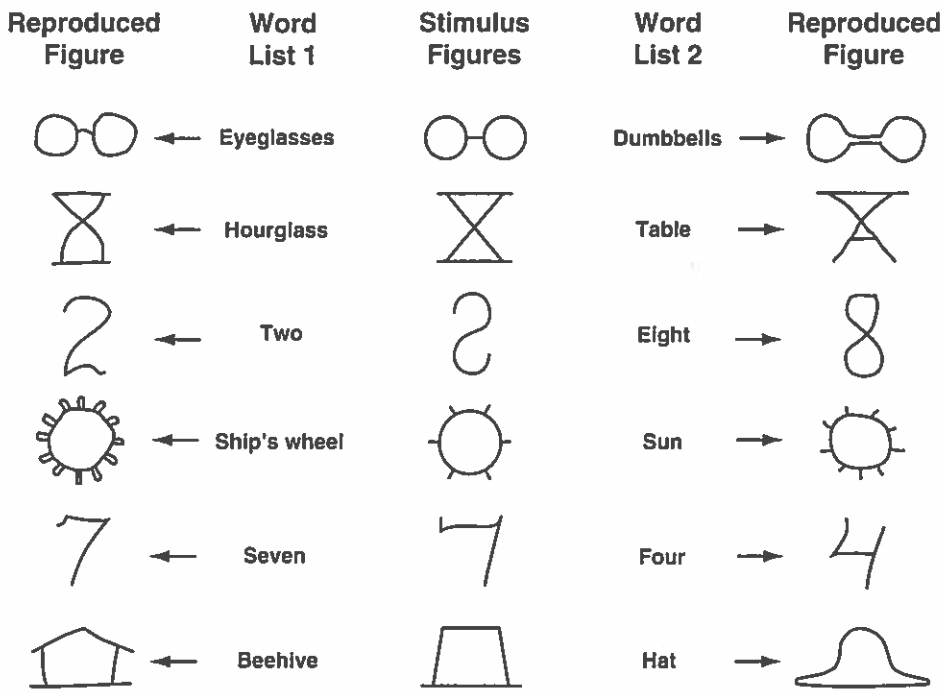

Figure 5.6 Several examples of the types of items presented by Carmichael, Hogan, and Walter. SOURCE: Carmichael, Hogan, & Walter (1932).

According to this view, a particular type of processing leads to better memory performance because it is appropriate for the kind of test that will be conducted. According to the related encoding specificity principle, the recollection of an event depends on the interaction between the properties of the encoded event and the properties of the retrieval information. Because it is the interaction that is important, no statements can be made about the mnemonic properties of an item, of a process, or of a cue. It is only when the interaction of the encoding and retrieval conditions is specified that meaningful statements about memory can be made. Processing approaches readily account for context-dependent (memory can be worse when tested in a new or different context) and context-congruent (a given context tends to cue memories that are consistent with that context) memory effects. Few if any tasks are process-pure. One way to separate the relative contributions of different processes is the process dissociation technique. This works by comparing performance in two conditions: one in which both processes work together and one in which they work in opposition.