

Handbook of Perception and Cognition
2nd Edition

Series Editors
Edward C. Carterette
and **Morton P. Friedman**

1996

Memory

Edited by
Elizabeth Ligon Bjork
Robert A. Bjork

Department of Psychology
University of California, Los Angeles
Los Angeles, California

1996



Academic Press

San Diego London Boston
New York Sydney Tokyo Toronto

Retrieval Processes

Henry L. Roediger, III
Melissa J. Guynn

The key process of memory is retrieval.
—Endel Tulving (1991)

I. INTRODUCTION

Processes of learning and memory are typically conceptualized as involving at least three stages: encoding, storage, and retrieval (Melton, 1963). *Encoding* refers to initial learning or acquisition of information. Within the context of chapters in the present volume, encoded information would certainly have passed through sensory storage systems and in most treatments would also have been recoded into a short-term store (or primary memory, or working memory). *Storage* refers to maintaining information over time. The usual description of the information during the storage stage is in terms of *memory traces*. The idea is that encoding processes leave a residue in the nervous system—memory traces—that persist over time. The third stage, *retrieval*, refers to accessing stored information. Retrieval processes, then, refer to the means of using stored information; they are the topic of this chapter. However, as the foregoing should make clear, retrieval processes are inextricably bound to those of encoding and storage. Although we can focus on factors that are manipulated at the retrieval stage, we cannot ignore the other stages. How an event is encoded and stored determines how well it can be retrieved later, and what cues will effect its retrieval.

It is a curious fact that the retrieval phase of memory was largely neglected by researchers until the mid-1960s. One could find occasional mention of its importance by experimental psychologists (Köhler, 1947; Melton,

1963), but investigators rarely created conditions to study retrieval directly. An unspoken assumption guiding much research was that responses produced on a memory test indicated the contents of storage. Failures to produce information were ascribed to problems in encoding or storage. For example, forgotten information might not have been transferred from a short-term store to a long-term store (Waugh & Norman, 1965), or it may not have been processed to a deep, semantic level (Craik & Lockhart, 1972), or over time memory traces may have decayed or been overwritten.

Tulving (1974) referred to theories, such as the original levels of processing ideas, as trace-dependent theories of memory: the critical determinant of performance on memory tests was thought to be the status of the memory trace at the time of testing. Little attention was paid to the idea that encoding and storage could have succeeded—the memory trace of the experience was robust—and that forgetting could still occur due to retrieval failure. The alternative to trace-dependent forgetting theories are those embodying the assumption of cue-dependent forgetting. Cue-dependent forgetting theorists maintain that “memory for an event is always a product of information from two sources,” the memory trace and the retrieval cue, the latter being “the information present in the individual’s cognitive environment when retrieval occurs” (Tulving, 1974, p. 74). The key idea is that *both* the traces of past experience and the information or cues in the cognitive system during the test are critical determinants of remembering.

Although the retrieval phase of the learning and memory process has generally been neglected by psychologists until the past thirty years, early in the century Richard Semon, a German scholar, advanced ideas on the critical role of retrieval processes in understanding memory (see Schacter, Eich, & Tulving, 1978, for a synopsis of Semon’s thought). Semon coined the name *engram* for memory trace and it caught on and was widely used. He also advanced the term *ecphory*, defined as “the influences which awaken the mnemonic trace or engram out of its latent state into one of manifested activity” (Semon, 1921, p. 12). *Ecphory* is roughly the same as *retrieval* and in propounding The Law of Ecphory as one of two general laws of memory, he noted that successful ecphory requires “the partial return” of the original encoding episode (as we might put it today).

Semon’s ideas were far ahead of their time and probably have not been completely mined by psychologists, even today. Tulving (1983) has championed use of Semon’s terms for retrieval processes, but they have not been widely accepted. Schacter (1982) has written a fascinating book on the life and work of Richard Semon and speculates about why his work did not have more impact at the time.

The interesting point of this story for present purposes is that the importance of retrieval processes was noted quite early in this century, yet it did not become the explicit object of study until much later (Tulving & Pearl-

stone, 1966; Weiner, 1966). However, by 1980, when Loftus and Loftus (1980) published a survey of psychologists interested in learning and memory, most of them said they believed that retrieval failures accounted for most cases of forgetting. Evidence presented in this chapter shows why so many psychologists came to believe that retrieval processes are critical determinants of recollection.

The purpose of the present chapter is to review what we know about retrieval processes in human memory. In Section II we review two basic ways of studying retrieval processes: giving retrieval cues during a test, and testing people repeatedly with the same cues. In Section III we describe one general principle that has been repeatedly emphasized as governing retrieval of memories and consider two other principles that seem to apply. We argue that understanding this small set of principles can provide considerable power in analyzing retrieval processes. In Section IV we introduce the encoding/retrieval paradigm (Tulving, 1983) and argue that it represents a fundamentally important method of studying retrieval processes and their interaction with encoding processes. In Sections IV–VI we (1) review how the encoding/retrieval paradigm has been applied to understand various phenomena; (2) discuss the effects of prior retrieval on later retrieval; and (3) briefly describe some related topics. Finally, we summarize the chapter's main points.

II. METHODS OF STUDYING RETRIEVAL

Psychologists interested in learning and memory have developed several distinct methods of studying retrieval. Indeed, in some sense, all methods of testing memory study retrieval, because a retrieval phase is involved in every memory test. However, we can single out some methods for special attention, given here or elsewhere in this volume. These are: (1) repeated testing (without intervening study opportunities); (2) presentation of cues at test; (3) judgments made during retrieval; and (4) comparison of different instructions at retrieval. In this chapter we cover results based on the first two techniques, repeated testing and presentation of retrieval cues (with greater emphasis on the latter). The method of requesting subjects to make judgments at test is covered by Metcalfe in Chapter 11 (about metacognitive processes). The use of different instructions at retrieval, as in Jacoby's (1991) process dissociation procedure or in the large number of techniques used to study implicit memory, is covered by Kelley and Lindsay in Chapter 2.

A. Repeated Testing: Reminiscence and Hypermnnesia

One venerable method for studying the role of retrieval processes is to test subjects' memories repeatedly, without opportunities for intervening study. W. Brown (1923) conducted some of the first well-known experiments. He

asked the fundamental question that formed the title of his paper, “To what extent is memory measured by a single recall trial?” Brown conducted two experiments to answer the question. In the first, he asked college students to recall as many of the 48 states as possible during a 5-min period. Half an hour later, without any warning, he gave them the same task again. The second experiment was similar, except in this case students were given 48 words to learn, with each word presented four times. Once again, the test involved an initial 5-min recall period immediately after learning and then a second test 30 min later.

The results from W. Brown’s (1923) experiments are presented in Table 1. The first two rows indicate the total number of items recalled on the first and the second test, respectively, with the difference between them ($T_2 - T_1$) in the third row. Brown discovered that subjects in both experiments, with either states or words in a list, recalled more items on the second test than on the first test. This increase in recall between tests is now called *hypernesia* (Erdelyi & Becker, 1974). The finding is a surprise in some ways because the second test occurred a considerable amount of time after the first test. Thus, one would normally expect forgetting to occur between the two tests, at least in the case of list learning. Indeed, McGeoch (1932) used Brown’s results as one prong in his attack on the law of disuse, or the decay theory of memory. How could memories decay with the passage of time, if more were recalled on a later test than on an earlier test?

W. Brown (1923) broke down performance into two components of interest in repeated testing, *intertest forgetting* and *intertest recovery*. Despite the fact that more items were recalled, overall, on the second test than on the first test, some items were actually forgotten between the two tests. That is, states or words recalled on the first test were forgotten on the second test, as shown in the fourth row of Table 1 (1.94 states and 3.04 words). The last row of the table shows the number of items recovered between the two tests, or items not recalled on the first test but recalled on the second test (5.29 states and 4.33 words). Obviously, the overall improvement between tests for states (and to a lesser extent for words) reflects the fact that *intertest recovery* exceeded *intertest forgetting*. This recovery component is often called *reminiscence*, after Ballard’s (1913) definition: “the remembering again of the forgotten without re-learning” (p. v).

W. Brown’s (1923) basic observations have been replicated many times. The results point out several basic facts about retrieval processes that will be repeatedly encountered in this chapter. A single test of memory is an imperfect indicator of knowledge. Give even the same test a few minutes later, and subjects will exhibit a different pattern of recall (sometimes remembering more, sometimes less—and different events will be recalled on the two occasions). Since the early experiments by Ballard (1913) and W. Brown (1923), many researchers have been intrigued by the observation that recall can improve across tests without intervening study opportunities. Reference

TABLE 1 Results from W. Brown's (1923) Experiments Using Repeated Testing

	States	Word list	
Test 1	36.31	25.48	
Test 2	39.66	26.77	
Difference ($T_2 - T_1$)	3.35	1.29	Hypermnesia
Interrest forgetting	1.94	3.04	
Interrest recovery	5.29	4.33	Reminiscence

experiments by Tulving (1967) and Erdelyi and Becker (1974) impressed upon modern researchers the need to study reminiscence and hypermnesia, respectively. Throughout the 1970s and 1980s, many researchers were occupied with these topics, and summaries may be found in reviews by Erdelyi (1984), Payne (1987), and Roediger and Challis (1989).

Research on reminiscence and hypermnesia points out a fact that is sometimes overlooked: retrieval of a set of material extends over a period of time. The practice of many researchers is to give a fixed, often quite short, time period for a test. However, in experiments in which retrieval time has been extended up to 21 min, recall continues to improve, albeit at a negatively accelerated pace (Roediger & Thorpe, 1978). Curiously, the study of how recall changes over time under different experimental conditions has not been extensively studied and may hold considerable promise as a future area of inquiry. Wixted and Rohrer (1994) provide an overview of cumulative measures of recall and how they can be analyzed.

In attempting to answer the question of why recall improves across tests, Roediger and Thorpe (1978) suggested that people use subjective retrieval cues, or cues they generate themselves, to guide retrieval. That is, even though the task is nominally one of free recall, in which no cues are provided, subjects retrieve information by providing themselves with cues (Tulving, 1974; see also Conway, this volume, Chapter 6). The general idea is similar to Estes' (1955) stimulus sampling theory of spontaneous recovery in animal conditioning and has been embedded in more formal models by Mensink and Raaijmakers (1988). However, almost by definition, it is impossible to manipulate the nature of the cues used in free recall and therefore the theories can receive only indirect support. Perhaps for this reason more researchers have been interested in studying retrieval processes through manipulation of overt retrieval cues, in what are known as cued recall paradigms.

B. Cued Recall Paradigms

The logic behind cued recall techniques to study retrieval is straightforward and powerful. In the simplest case, two groups of subjects are exposed to

the same set of material under identical study conditions and are treated in exactly the same way until the moment they are tested. Therefore, encoding and storage of the material can be assumed to be the same in the two groups. At the time of the test, the retrieval conditions of the two groups are manipulated in a systematic way (such as through the instructions given, the physical environment in which testing takes place, or the overt cues provided). Any differences in retention can then be attributed to differences in the retrieval processes invoked by the test conditions. Tulving and Pearlstone (1966) were the first to make this logic explicit and to provide an experiment to demonstrate it. We present only selected conditions to make their essential points.

Tulving and Pearlstone (1966) presented high school students with categorized word lists (items belonging to common categories such as articles of clothing, automobiles, or types of birds). In one of the study conditions, subjects were given lists of 48 words, with 2 words in each of 24 categories. Specifically, subjects heard lists such as "articles of clothing: blouse, sweater; types of birds: blue jay, parakeet," and so forth. Subjects were told to remember the 48 list words for an unspecified test.

Two test conditions were of interest. One was free recall, in which subjects were given a blank sheet of paper and asked to recall as many words as possible from the list. On average, they recalled 19.3 words correctly. We can then ask what happened to the other 29 or so? Were they not acquired? Not stored? Or was the failure one of retrieval? Absolute answers to these questions are not possible, but a companion condition in the Tulving and Pearlstone (1966) experiment indicates that retrieval difficulties likely were responsible for much of the forgetting. A second group of subjects, who had been given the same list of 48 categorized words, received a cued recall test in which they were presented with the 24 category names as recall cues. In this condition, subjects recalled 35.9 words. Therefore, with all encoding and storage conditions held constant, presentation of retrieval cues almost doubled the number of events recalled. This difference indicates that, for the free-recall subjects, information was *available* (i.e., stored) that was not *accessible* (i.e., retrievable) on a free-recall test. Tulving and Pearlstone emphasized this distinction between availability and accessibility of memories. Psychologists would like to have a method that could accurately tell what information is stored in memory, but such a technique is impossible. All we can ever know is what information is accessible under a particular set of retrieval conditions.

By the way, we should mention that an artifactual explanation of the Tulving and Pearlstone (1966) result in terms of guessing does not work. They had eliminated from consideration the most obvious associates to the category names and used other means to protect against guessing. Interestingly, in most cued and free-recall situations, subjects seem to guess very

little unless instructed to do so. M. J. Watkins and Gardiner (1982) have discussed various precautions against guessing in cued recall experiments.

The basic procedure in providing people with specific cues to prompt or aid retrieval has become the dominant method of studying retrieval processes. The remainder of the chapter discusses results that have been obtained using this technique. First, however, we discuss some general principles that have guided research in this area.

III. PRINCIPLES GOVERNING RETRIEVAL

In his famous 1932 article attacking the law of disuse, McGeoch argued that two conditions increased forgetting. One was retroactive interference, which was heavily studied in the thirty years after his paper (see Anderson & Neely, this volume, Chapter 8, for a discussion of interference effects in memory). The other principle, mentioned only in passing, was “altered stimulating conditions” between learning and testing. The implication is that the opposite of such altered conditions would promote remembering: to the extent that the similarity between “stimulating conditions” during learning and testing was increased, good retention should result.

Even though McGeoch (1932) considered the stimulating conditions to be one of the critical factors, no one paid much experimental attention to this topic for over thirty years. Perhaps the reason is that the principle seemed obvious. After all, experimental psychologists of this era followed the animal conditioning literature closely, and McGeoch’s principle seemed quite close to that of stimulus generalization, first noted by Pavlov (1927). In either classical or operant conditioning situations, if an animal is trained to respond in the presence of a particular stimulus, the closer a test stimulus is to that original training stimulus on some physical dimension, the greater is the responding. This principle of primary stimulus generalization may have been so ingrained in the thinking of experimental psychologists that they failed to investigate similar principles in human learning and memory, with but few exceptions (Pan, 1926; see Murdock, 1989, for discussion).

It was not until the late 1960s, when Tulving and colleagues announced the encoding specificity principle (Tulving & Osler, 1968; Tulving & Thomson, 1973), that interest in the match between properties of cues at test and those of “stimulating conditions” during study was aroused. As Tulving (1983, p. 223) put it, “. . . recollection of an event, or a certain aspect of it, occurs if and only if properties of the trace of the event are sufficiently similar to the retrieval information” provided in the retrieval cues. As we describe later, several lines of research can be interpreted under the broad framework of the encoding specificity hypothesis. Most of the research conducted within this framework has emphasized quite specific cues and how they matched (or did not match) the encoding of studied events.

A broadening of the encoding specificity ideas occurred in the notion of transfer-appropriate processing, endorsed by Bransford, Franks, Morris, and Stein (1979). The basic idea is that performance on a memory test can be considered a case of transfer, and that transfer will be affected by the match of processing activities during encoding and retrieval. The greater the match between encoding activities and retrieval activities, the greater positive transfer there will be. The emphasis is on mental processes and is similar to that embedded in Kolers' (1979) procedural approach to memory (see also Kolers & Roediger, 1984).

Roughly the same central idea is captured in the ideas of altered stimulating conditions, stimulus generalization, the encoding specificity principle, and transfer-appropriate processing. Maximizing the similarity (in overt stimulus conditions, in specific encoding, or in processing activities) between a study and a test occasion benefits retention. This general principle has guided much research in learning and memory and has been used to organize much of the rest of this chapter. However, there is one major difference between the encoding specificity and transfer-appropriate processing principles and the previous ideas. The ideas of stimulus generalization and altered stimulating conditions (or generalization decrement) focus completely on the external stimulus conditions, whereas encoding specificity emphasizes how the stimulus is coded. According to the latter idea, the properties of the stimulus *as coded* determine retention. For example, in some cases a cue that is different from the original studied event can provoke its recollection better than a literal copy of the event itself (Tulving & Thomson, 1973). Therefore, physical similarity of study and test events is not the crucial determinant of retention; rather, psychological similarity—the similarity of encoding and retrieval processes—is critical.

Other principles also are useful in understanding retrieval processes. One is the cue overload principle noted by Earhard (1967) and championed by O. C. Watkins and M. J. Watkins (1975; see also M. J. Watkins, 1979). The basic idea of this principle is that the more information that is subsumed under a single retrieval cue, the poorer will be the recall for any one piece of information. Consider Tulving and Pearlstone's (1966) situation using category name retrieval cues. The more items presented in a given category, the less likely that category name is to cue any particular item. The cue is said to be overloaded. For example, Roediger (1973) presented subjects with categorized lists with either four, five, six, or seven items per category. Probabilities of recalling words from a category, given its name as a cue, were .69, .65, .64, and .59 as the category size increased. Doubtless, probability of recall would have been even higher with only one or two items per category.

Another related principle governing retrieval is that of distinctiveness:

Events that are distinctive, or that stand out from a background of other similar events by being different, are generally well remembered. For example, you can probably remember the events you experienced on your most recent birthday better than you can the events that occurred either a few days before or a few days after your birthday. Unusual or distinctive events may be well remembered because they represent powerful retrieval cues that are not overloaded (i.e., relatively few events are subsumed under them, compared to other types of cues). We discuss this principle more fully later in the chapter.

Such general principles that govern retrieval may not be satisfying to those who seek molecular models to account for memory performance, but for purposes of a chapter reviewing a considerable body of research, this approach seems defensible. We turn next to the primary paradigm that has been used to study retrieval processes and their interaction with encoding conditions.

IV. THE ENCODING/RETRIEVAL PARADIGM

Consider the hypothetical arrangement of conditions in Figure 1. There are two encoding conditions (A and B) and two retrieval conditions (A' and B'). If we consider first the rows, and only one of the columns, we have an *encoding experiment*. That is, experimenters manipulate encoding strategies or material and examine performance on a single memory test. For example, the encoding dimension might be rehearsal versus imagery instructions, superficial versus meaningful processing requirements, pictures versus words, or high-imagery words versus low-imagery words. The important point is that encoding conditions are manipulated and test conditions are held constant. Consider next performance in the columns and only one of the rows. In this case, encoding conditions are held constant while test conditions are manipulated. This is a *retrieval experiment*. An example is the Tulving and Pearlstone (1966) experiment described earlier in which subjects studied a list of words and were then tested with either free recall or cued recall.

Both encoding experiments (the columns) and retrieval experiments (the rows) are quite common in the experimental study of memory and have yielded much useful information. However, simultaneous manipulation of both encoding and retrieval conditions permits a more powerful design, referred to by Tulving (1983) as the encoding/retrieval paradigm. Briefly, in this abstract case, we consider encoding conditions A and B to differ on some dimension, and retrieval conditions A' and B' to be similar on this dimension to encoding conditions A and B, respectively. If the principle of encoding specificity or transfer-appropriate processing holds, performance

		Retrieval Condition	
		A'	B'
Encoding Condition	A	A-A'	A-B'
	B	B-A'	B-B'

FIGURE 1 The encoding/retrieval paradigm. Minimally, two encoding conditions (A and B) are crossed with two retrieval conditions (A' and B'), often designed to be similar in some way to the encoding conditions. Retention should be enhanced in conditions represented by cells in which the best match exists between study and test conditions (A-A', B-B') relative to the other conditions (A-B', B-A'). (After Tulving, 1983, p. 220, with permission of Oxford University Press.)

in conditions A-A' and B-B' (where encoding and retrieval conditions match) should be better than performance in conditions A-B' and B-A' (where encoding and retrieval conditions match less well).

This basic paradigm has been used to investigate several problems in the psychology of memory and points out the critical interaction of encoding and retrieval processes. We consider below the following topics: (1) manipulating verbal context at study and test; (2) manipulating physical context or environmental context (locations) at study and test; (3) manipulating the subjects' internal context (their drug state or mood state) at study and test; (4) manipulating mental and physical operations applied to material at study and test; and (5) manipulating the type of information emphasized at study and test (relational or item specific). In each case, performance depends critically on the match of processes or information between study and test occasions.

A. Manipulation of Verbal Context

Much research on the effects of verbal context on remembering has employed the encoding/retrieval paradigm. People study the same nominal target (a word) in one or another context. Then, when they are tested later, cues are presented that instantiate either the same context as at study or a different context. Among the first such experiments were those reported by Tulving and Osler (1968) and Thomson and Tulving (1970). The authors were asking the critical question: Under what conditions is a retrieval cue effective? For example, prior research had shown that if people study a word like *black* in a list of otherwise unrelated words, they are much more likely to recall *black* in the context of a strong associate like *white* than they

are under conditions of free recall (Bilodeau & Blick, 1965). Why are these strong associates effective retrieval cues?

Thomson and Tulving (1970, Experiment 2) presented subjects with target words such as *BLACK* with either no context word at study or with a weakly associated word (*train-BLACK*) as context at study. Subjects were told that their task was to remember the capitalized word, but they should also note its relation to the context word (when it was present) to aid later recall. These two study conditions were crossed with three test conditions: Subjects got either no cues (free recall), or the weak associates as cues (*train-???*), or the strong associates as cues (*white-???*).

Shown in Table 2 are the results of the experiment. Examine first the top row, which presents performance for words studied with no context cues. Under conditions of free recall, people could remember 49% of 24 target words; however, with strong associates as retrieval cues, they recalled reliably more, 68%. On the other hand, weak associates given at the time of test did not aid recall; indeed, these irrelevant cues seemed to hurt recall a bit (43% recalled). Now consider the second row, which presents performance for the same target words studied in the context of weakly associated words. Free recall of these words suffered relative to the no-context (at study) condition (after all, subjects studied twice as many words), but the real interest is in how the effectiveness of the weak and strong associate cues changed. Now, recall was quite good with weak associate cues (82%), but quite poor—even worse than in free recall—with the strong associate cues (23%)!

The results of Thomson and Tulving's (1970) experiment, combined with the earlier ones of Tulving and Osler (1968), led to the encoding specificity hypothesis as the principle governing the effectiveness of retrieval cues: A retrieval cue will be effective if and only if it reinstates the original encoding of the to-be-remembered event. When a word like *black* is presented without context, it is presumably encoded with regard to its dominant meaning (as associated with *white*). Therefore, *white* serves as an effective retrieval cue, and a weak associate like *train* does not. However,

TABLE 2 Results of Thomson and Tulving's Experiment 2 (1970)^a

Study context	Test context/cues		
	No cues	Weak associates	Strong associates
No cues (Black)	.49	.43	.68
Weak associates (train-Black)	.30	.82	.23

^aSubjects studied 24 target words (*BLACK*) without context or in the context of a weak associate (*train-BLACK*). Then they were provided with various test contexts: no cues (free recall), weakly associated cues (*train-???*), or strongly associated cues (*white-???*). Results are the proportion of words recalled in the various conditions. See text for explanation.

when *black* is encoded in the context of a weak associate like *train*, subjects are likely to engage in a more idiosyncratic encoding of the target word (e.g., they might imagine a black train). In this case, the weak associate could serve as an excellent retrieval cue, but now the strong associate is completely ineffective, even relative to free recall. So even a strong associate will not be an effective retrieval cue for a word if the study context led subjects to encode the word in a nonstandard manner.

One surprising aspect of Thomson and Tulving's (1970) experiment was that free recall could be superior to cued recall with strong associates as cues. Some later studies by Newman and Frith (1977), among others, questioned this finding, but more recent results have confirmed it and provided further support for the encoding specificity hypothesis (Roediger & Adelson, 1980; Roediger & Payne, 1983).

Another impressive demonstration of the role of encoding factors and the effectiveness of retrieval cues was provided by Barclay, Bransford, Franks, McCarrell, and Nitsch (1974). Their experiment used sentences rather than single words, although the target for recall was always a word in the sentence. For example, subjects studied sentences such as "The man lifted the piano" or "The man tuned the piano." A retrieval cue such as "something heavy" was predicted to promote better recall for the first sentence, whereas the cue "something with a nice sound" was predicted to promote better recall for the second sentence. The results bore out these predictions, as 47% of the target words were recalled to appropriate cues, but only 16% to inappropriate cues. As with other results of this ilk, their findings show that words are semantically flexible; depending on the context, different features of the word will be encoded and therefore different cues will be effective, even for the same nominal target (*piano* in this example).

The role of verbal context has been shown in dozens—and maybe even a hundred—experiments. However, the great bulk of these experiments make the same point: Matching of the verbal context between study and test in which a target event appears greatly determines its memorability. The "matching" in these cases is usually on some meaningful dimension (Roediger & Adelson, 1980). However, on other types of memory tests—perceptual implicit memory tests—manipulations of meaning play little role, and instead, matching of perceptual features between study and test produces better performance (see Kelley & Lindsay, this volume, Chapter 2; Roediger, 1990). However, for most explicit memory tests, or those invoking conscious recollection, matching of meaning of events between study and test largely determines their memorability.

B. Manipulation of Physical Context

Each of us has had the experience of returning to a place from some past time in our lives and being suddenly reminded of many of the events and

people of that time. The place may be an elementary school, a former residence, or a city from which one has moved away. The experience is so compelling that we easily come to believe that location (and other physical features) serve as potent retrieval cues reminding us of these distant events. But is it the case that these physical cues actually enable us to remember more than we could without them, as compared, for example, to someone merely asking us to try to recall events from that particular time and place? That is, do the actual physical locations serve as better retrieval cues than would verbal cues designed to help us retrieve the original event? The anecdotes that we can all recount about sudden recollection after returning to familiar places do not answer this question.

Numerous experiments have been conducted in which subjects learn material in one of two distinct physical contexts and then are tested on it either in the same context or in a different context. Such experiments conform nicely to the encoding/retrieval paradigm. If physical context modulates memory, one would expect better retention when testing occurs in the same place as studying rather than in a different place. In the mid-1970s, two spectacular examples of such interactions were reported. Godden and Baddeley (1975) conducted an experiment in which skin divers heard a list of 36 words in one of two different contexts—under water or on land—and were later tested for recall in the same physical environment in which the list had been studied or in the other environment. The subjects wore wet suits in both contexts and recalled the words orally. The results are shown in Table 3, where it can be seen that a greater percentage of words was recalled when the context at study and test matched than when it did not match.

A second striking series of experiments reporting context dependency were conducted by S. M. Smith, Glenberg, and Bjork (1978). In their Experiment 3, subjects studied 80 words (from 10 categories, with 8 items per category) in one of two different rooms. The rooms were made distinctive on several dimensions (size, classroom or storeroom, the experimenter, equipment, scent in the room, time of day for the test) and subjects learned

TABLE 3 Results of Godden and Baddeley (1975)^a

Study context	Test context	
	Land	Water
Land	.38	.24
Water	.23	.32

^aSubjects studied 36 words on land or under water and recalled them in the same context or in the other context. The proportion of words recalled was greater when the test context matched the study context.

the list either in the same room in which they were tested or in the other room. Subjects were given a free-recall test 24 hr later. When subjects were tested in the same room in which they had studied, they averaged about 49 words correct; when they were tested in a different room, they averaged only 35 words correct. Once again, the results showed striking specificity of matching encoding and retrieval contexts. In a later experiment, S. M. Smith (1979, Experiment 2) showed that most of the disadvantage of being tested in a different context from the learning context could be eliminated if subjects imagined being in the room in which they had learned the material while taking the test.

Curiously, despite these impressive findings from the mid-1970s, the effects of physical context on memory have been difficult to replicate. Although there have been some successes, they have also been numerous failures. Fernandez and Glenberg (1985) conducted eight experiments examining context (room) dependency, varying numerous factors across the experiments, but in none of the experiments did an effect of context dependency appear. In one case, the experiment was conducted in even the same locations used by S. M. Smith et al. (1978), so the failure to replicate is especially puzzling. Averaging over all experiments (weighting each equally), the probability of recalling words in the same context was .29 and in the different context was .30.

Saufley, Otaka, and Bavaresco (1985) capitalized on a naturally occurring context manipulation present on many university campuses. Often at large universities, students hear lectures in large halls and then are assigned to take their examinations either in the same lecture hall or in small ancillary classrooms. Saufley et al. tested Introductory Psychology students over three years at the University of California at Berkeley in just such a situation, where students were randomly assigned to take their exams either in the main lecture hall (where they had been taught) or in smaller overflow rooms. Across 21 "experiments" in this classroom setting, they failed to find any difference between the performance of the 3613 students who took the test in the same room and the 2412 who took the test in a different room. Subjects recognized, on average, about 67% of the material in each context. Although this outcome may be comforting for students who might find themselves in this circumstance, the failure to find any effect of room dependency is disquieting, given the prior literature. However, all tests taken by students were multiple choice (recognition memory) tests, not free-recall tests, and, as we cover in the next section, sometimes encoding/retrieval interactions occur on free-recall tests but not on recognition tests.

Why does environmental context have such a puzzling relation to memory retrieval? Why do some researchers obtain striking effects and others none at all? Bjork and Richardson-Klavehn (1989) and S. M. Smith (1988)

have considered a number of possible reasons. Some differences between the naturally occurring cases and the laboratory experiments are quite obvious. For one, many researchers use only very short retention intervals between study and test of material, whereas the most compelling examples from our lives occur when the retention interval is very great (when we return to a place from which we have been absent for years). In addition, the type of materials, conditions of learning, and many other factors differ between the naturally occurring cases and the laboratory experiments. These factors may explain differences between our experiences and laboratory experiments, but they certainly do not explain failures to replicate across relatively similar experiments.

Another possibility is suggested from experiments reported by Eich (1985). He had subjects study lists of 24 words in one of two rooms. Different groups of subjects were told to use different types of imagery during encoding. In one condition (integrated imagery), subjects were asked to generate an image in which the named object was seen as being conjoined with a particular feature in their environment (e.g., "I imagine a diamond-shaped *kite* lying on top of the table located in the corner of the room."). Other subjects were told to form isolated images in which the object was visualized as existing by itself (e.g., "I imagine a big blue *kite* sailing in the sky."). They were to rate the vividness of their images on a scale of 1 to 7 in both conditions. Subjects were tested 48 hr later, first on a free-recall test and then on a two-alternative forced-choice recognition test, and either in the same room or in a different room.

The results are shown in Panel A of Table 4, with free recall in the columns on the left and recognition in the columns on the right. We consider the recall results first. As can be seen, an effect of room context was obtained in free recall, but only in the case when integrated images were

TABLE 4 Results of Eich (1985)^{a,b}

Imagery condition	Free recall		Recognition	
	Context		Context	
	Same	Different	Same	Different
A Isolated	.26	.24	.95	.93
A Integrated	.45	.31	.90	.91
B Isolated	Free recall		Recognition	
B Integrated	.25		.94	
	.38		.90	

^aResults in Panel B are collapsed across the Same and Different contexts.

^bSubjects studied 24 words in one room and were tested on recall and on recognition in the same room or in a different room.

formed during the study phase (in the second row). Subjects recalled 45% when tested in the same room and 31% when tested in a different room. Presumably subjects associated their images of the target words to items in the room and then later these items served as effective retrieval cues when subjects were tested in the same room, which of course they could not do when subjects were tested in a different room. When studied events were imaged in isolation, a room context effect did not occur, with about 25% recall in each condition (in the first row). Also, note that, overall, subjects recalled words better in the integrated imagery condition than in the isolated imagery condition (see Panel B, on the left).

One possibility, then, is that the mixed evidence with regard to the effects of physical context on recall is due to subtle factors that determine whether studied items are associated with features of the room. Subjects may need to employ conscious strategies to relate target items to features of the room for the physical context to serve as a useful cue later. If this association occurs, then maintaining the same context or changing to a different context will affect performance. If associations are not made between the recall targets and features of the context, then maintaining the context or changing it will not affect performance (as in Eich's, 1985, isolated imagery condition). Although this interpretation is plausible, it is still unsatisfying in that there is no good reason to believe that instructions were actually varied in subtle ways across previous experiments in the literature in a way that would explain successes and failures to find context-dependency effects.

The recognition results appear on the right of Table 4. Although their interpretation is clouded by an apparent ceiling effect—subjects were very near perfect performance of 1.0—two other interesting findings emerge. First, in the recognition test, no hint of a context-dependency effect occurred in either study condition (Panel A). The second finding can be seen in Panel B, where results have been collapsed across the same and different context conditions. Now the experiment appears as an encoding/retrieval experiment in which isolated and integrated imagery were the study conditions and free recall and recognition were the test conditions. Interestingly, it can be seen that the integrated imagery condition led to better performance on the free-recall test than did the isolated imagery condition, but that the opposite effect occurred on the recognition test; that is, although performance is near the ceiling, the isolated imagery condition led to significantly better recognition than did the integrated imagery condition. Note that this outcome occurred despite the fact that the free-recall test occurred prior to the recognition test and probably differentially boosted recognition in the integrated imagery condition (i.e., recall was greater in the integrated imagery condition and the act of recall probably boosted recognition). The finding of an independent variable having opposite effects on free recall and

recognition indicates that different processes operate in the two tests. We consider other results like these later in the section on effects of item-specific and relational encoding.

To return to the main question of why some experiments show effects of room context and others do not, perhaps the answer lies in subtle aspects of experimental instructions or manipulations that encourage subjects in some experiments to code the to-be-remembered items with regard to the room context, whereas others do not. These conditions would correspond to the integrated imagery and isolated imagery instructions, respectively, in Eich's (1985) experiment. Of course, Eich's experiment is only suggestive with regard to this point and by no means proves that this difference caused the discrepancies in the literature. However, at least Eich has identified a plausible candidate to explain these differences, which is a first step. McDaniel, Anderson, Einstein, and O'Halloran (1989) have alternatively proposed that if the original encoding was rich enough, subjects would be able to supply their own retrieval cues during the test and, accordingly, would not need to rely on environmental context. With weaker encoding strategies, however, greater reliance on environmental context at retrieval would lead to benefits of matching context at encoding and retrieval. S. M. Smith (1988) has proposed a similar idea. Despite the evidence from Eich (1985) and from McDaniel, Anderson, Einstein, and O'Halloran (1989), the effects of environmental context remain a puzzle.

One other experiment deserves mention before we leave this section, as it provides a convenient transition to the next topic. Schab (1990) asked whether presentation of a distinctive odor during learning and testing would aid memory. He presented adjectives under incidental learning conditions. During both the learning session and a later free-recall test, subjects were told to imagine the smell of chocolate, but in some cases they actually received a chocolate smell infused into the room and in other cases they did not. As shown in Table 5, when subjects experienced the chocolate odor during both the study phase and the test phase, their free-recall performance

TABLE 5 Results of Schab's Experiment 1 (1990)^a

Encoding conditions	Retrieval conditions	
	Odor	No odor
Odor	.21	.17
No odor	.13	.14

^aSubjects learned a list and were tested on it with a distinctive odor occurring both during study and test, during one occasion but not the other, or on neither occasion.

was better than when the odor occurred only at study, only at test, or not at all. This pattern was also replicated in later experiments and therefore represents another effect, albeit a rather small one, of environmental context on memory. In the next section, we consider somewhat more radical transformations of the subjects' internal context, created by introducing drugs into their systems.

C. State-Dependent Retrieval

Research on state-dependent retrieval involves manipulating the subject's internal state, usually in the form of changing the pharmacological state or the mood state of the individual. The literature on state-dependent retrieval is complex, but, at least in the case of manipulating drug states, some clarification has been gained. The business of how mood affects memory, however, is still rather uncertain.

The study of state-dependent retrieval arose from clinical observations. Psychologists and psychiatrists working with alcoholics occasionally noticed a curious phenomenon. Alcoholics would perform some act while under the influence of alcohol (e.g., hide a paycheck) and then, after becoming sober, forget where the paycheck was hidden. Of course, such alcoholic amnesia is not in itself surprising, because alcohol is known to have strong detrimental effects on memory and other cognitive processes. The curious event was that when the alcoholic fell off the wagon again, he could remember where the paycheck was hidden.

How can this phenomenon be explained? In line with the general principles of this chapter, the pharmacological state of the individual may serve as a retrieval cue; when intoxicated during an experience, the person will be better able to retrieve it later when intoxicated again than when sober. The surprising prediction is that retrieval under conditions of intoxication might actually surpass that in a sober state, at least when original learning occurred under conditions of intoxication. Retrieval depends on a match of the "state" of the person between study and test, in line with the general transfer-appropriate processing view of this chapter.

The early experimental literature on state-dependent retrieval was rather discouraging: Many experiments obtained the phenomenon (matching drug states at study and test led to better retention than mismatching states), but many experiments obtained null findings. Because of this disquieting state of affairs, many researchers concluded that state-dependent retrieval was not a reliable phenomenon. For example, Hilgard and Bower (1975, p. 547), in their influential textbook, stated that evidence for state-dependent retrieval in humans "rests on precarious grounds."

A seminal paper by Eich (1980; see also Eich, 1989) helped to straighten out the disparate findings. A hint to the solution was provided in an earlier

paper by Eich, Weingartner, Stillman, and Gillin (1975). With the permission of appropriate government agencies, they tested the state-dependent memory effects of marijuana. Subjects served in four different conditions depending on their state (drugged or sober at the time of learning and testing of the material). To induce the drug state, people smoked a marijuana cigarette 20 min before they learned the material or were tested on it. In the nondrugged state, they smoked a cigarette that tasted like marijuana but from which the active ingredient had been removed. To insure that the drug was having an effect, the experimenters checked both objective measures such as heart rate (marijuana increases it) and subjective reports obtained from rating scales (asking, e.g., "How 'high' do you feel?"). Marijuana had a large effect on both measures relative to the placebo condition.

The volunteers who served in the experiment were tested in the four conditions shown in Table 6. They studied lists of words in either the placebo or drug condition and then were tested in either the placebo or drug state 4 hr later. The lists consisted of 48 words from 24 different categories, with 2 words per category (as in the Tulving and Pearlstone, 1966, experiment described earlier). Two types of test were given: either free recall or category-name cued recall, also like the Tulving and Pearlstone study.

Results of their experiment are shown in Table 6, for both free recall and cued recall. First examine the free-recall results on the left. Not surprisingly, performance was best for people who were sober both when they studied the material and were tested on it (11.5 words recalled). If the subjects learned the material when sober but were tested when under the influence of marijuana, they did not recall as many words (9.9 words recalled). Therefore, as in other studies, marijuana (and by extension, other depressant drugs) impedes retrieval of information. The subjects did even worse if they studied the material under marijuana and then were tested sober (6.7 words recalled), consistent with the usual impairment of learning and memory by depressive drugs when people are removed from the drug.

In the critical condition for the state-dependent retrieval hypothesis—

TABLE 6 Results of the Eich, Weingartner, Stillman, & Gillin (1975) Experiment on State-Dependent Retrieval^a

Condition		Mean number of words recalled	
Study	Test	Free recall	Cued recall
Placebo	Placebo	11.5	24.0
Placebo	Drug	9.9	23.7
Drug	Placebo	6.7	22.6
Drug	Drug	10.5	22.3

^aSubjects received marijuana or a placebo before learning and testing of materials under free recall and cued recall conditions.

where people studied information in a drug state and then were tested in the same state 4 hr later—people recalled the material better than those who also had learned while under the drug but were tested sober. As can be seen in the Drug–Drug condition in Table 6, when people acquired material in a drug state, they recalled more words when placed back in the drug state (10.5) than when tested sober (6.7). Of course, this is no recommendation to take depressive drugs to help remember, because even in the Drug–Drug condition, people recalled fewer words than in the Placebo–Placebo condition, when they were sober at both times.

Eich et al.'s (1975) free-recall results impressively confirmed the concept of state-dependent retrieval. Note, however, the cued recall results on the right, where no evidence of state-dependent retrieval occurred. When given category name retrieval cues, people in the Drug–Drug condition recalled no more words than those tested in the Drug–Placebo condition, unlike the free recall results. Of course, cued recall produced better performance in all conditions than did free recall, but the provision of cues also eliminated the state-dependent retrieval effect.

Eich (1980) examined many experiments conducted on the state-dependent retrieval phenomenon. He discovered that only about half of all studies showed a reliable state-dependent effect (i.e., performance better in the drug–drug state than in the drug–placebo state). However, when he separated the experiments into groups based on the type of test given, he discovered that almost all experiments that had used a noncued (free recall or serial recall) test had found the state-dependent retrieval phenomenon, whereas very few experiments that had used cued recall tests had found the effect. This general conclusion—that state-dependent retrieval occurs under conditions of free recall but not cued recall—has generally held up over the intervening years (Eich, 1989). Drug state can serve as a retrieval cue, but more powerful overt cues such as category names can overshadow the weaker state cues and eliminate their effect. S. M. Smith (1988) argued for a similar principle operating in place-dependent retrieval experiments.

D. Mood-Dependent Retrieval

Drugs obviously affect the internal state of the organism. A similar manipulation involves changing people's moods. If people learn something while they are happy, will they be able to retrieve the information better if they are happy—relative to being in a neutral or sad mood—during the test? Does mood also serve as an internal contextual cue similar to drug states? Readers will probably not be surprised at this point to learn that the literature on this topic is mixed, with some researchers finding mood-dependent memory effects (Bower, 1981) and others not (see Blaney, 1986, for a review of findings). In most of these experiments, subjects are induced into happy or

sad moods by watching either comedy video tapes or sad or tragic video tapes, or by listening to uplifting music or dreary, depressing music. Typically, after the mood has been induced, subjects are presented with lists of words or other similar material to remember.

Eich and Metcalfe (1989) have reported an experiment that may help to unravel the puzzling effects of mood and memory. They had subjects either read words during study or generate them from a conceptual clue (e.g., either read *cold*, or generate it as the antonym of hot-??). Before the study phase, subjects listened to classical music that was designed to be either uplifting or depressing and they periodically rated their mood on a scale. Once subjects had met the experimenter's criterion for a mood change in the appropriate direction, the study material was presented. At a later point in time, subjects were reintroduced to the experimental setting, listened to classical music to induce either the same mood or the other mood, and then were given a free-recall test.

Eich and Metcalfe's (1989) results are shown in Table 7. The encoding/retrieval interaction for the read words, on the left, was not significant (although there was a slight tendency for recall in the matched mood cases to be a few percentage points higher than in the cases in which mood states were mismatched). However, the mood-dependency effect did appear, and in a striking fashion, for words that were generated during study, as shown in the data on the right of Table 7. Recall of generated words was almost twice as great when mood states matched between study and test as when they mismatched.

Eich and Metcalfe's (1989) results for the read words did not seem to represent a floor effect, because in later experiments they brought performance up by having subjects read words several times, and yet the mood-dependent effect still did not occur. The authors argued that mood-dependent effects might occur only for self-initiated activity on the part of the subjects, and not so powerfully for events presented externally. Subjects' moods may function as potent retrieval cues for their internally generated thoughts, but not for their externally produced reactions. Because most

TABLE 7 Results of Eich and Metcalfe (1989)^a

Encoding condition	Test condition			
	Read words		Generated words	
	Happy	Sad	Happy	Sad
Happy	.09	.04	.32	.17
Sad	.05	.07	.17	.27

^aSubjects read or generated words in a happy or sad mood and then were tested later in the same mood or the opposite mood.

prior experiments used external presentation of stimuli, perhaps mood-dependent effects were small and variable. This is an interesting hypothesis and its application to this field warrants further research. Eich, Macaulay, and Ryan (1994) provided additional evidence for its validity.

E. Manipulation of Mental and Physical Operations

Another means of implementing the encoding/retrieval paradigm is to require subjects to perform similar or different mental operations at encoding and retrieval. A classic experiment of this type reported by Morris, Bransford, and Franks (1977) introduced the transfer-appropriate processing approach. They were interested in how manipulations designed to influence the level of processing of studied stimuli affected performance on different types of memory tests. The usual expectation is that deeper, more meaningful processing should increase retention compared to more shallow or superficial processing (Craik & Lockhart, 1972; Craik & Tulving, 1975). Morris et al. (Experiment 1) had subjects study the same words in sentences that promoted either phonemic or semantic encoding of the words. For example, subjects studied words such as *Eagle* and answered questions like “_____ rhymes with legal” or “_____ is a large bird.” Subjects responded *yes* or *no* to each statement, and we consider results based on tests of items to which subjects responded *yes* during study.

The subjects’ memories were tested in two different ways, with the tests designed either to match or to mismatch the way the information was encoded. Half the subjects were tested on a standard recognition test in which studied words were intermixed with nonstudied words and the task was to identify the studied words. Morris et al. (1977) assumed that subjects accomplished this task by referring to the meaning of the test words, resulting in the typical levels-of-processing effect: better performance for words encoded semantically than phonemically. Indeed, as is apparent in the left column of Table 8, just this pattern occurred. The novel test condition involved a rhyme recognition test. Subjects were told that the test items included words that rhymed with studied words and that they should discriminate these rhyming words from distractors that did not rhyme with the targets. The results of this rhyme recognition test appear in the right column of Table 8, which shows that the standard levels-of-processing effect was reversed. Now phonemic encoding produced better performance than did semantic encoding, in general conformity with the principle of transfer-appropriate processing (or encoding specificity).

The Morris et al. (1977) experiment revealed two important lessons about retrieval processes that we have already met in other forms. First, even so powerful a variable as level of processing does not have a uniform effect across all types of test. On some tests, the levels-of-processing effect

TABLE 8 Results of Morris, Bransford, & Frank's Experiment 1 (1977)^a

Encoding condition	Test conditions	
	Standard recognition semantic	Rhyme recognition phonemic
Semantic	.84	.33
Phonemic	.63	.49

^aStudy conditions encouraged phonemic or semantic processing and test conditions were designed to tap the two types of knowledge.

disappears (Fisher & Craik, 1977; McDaniel, Friedman, & Bourne, 1978) or even reverses (as in the Morris et al. results). Second, matching the processing required by study and test conditions enhances performance. It should be noted, however, that there was still a main effect such that semantic study and test processing produced better overall performance than did phonemic study and test processing (the upper left cell compared to the lower right cell in Table 8). Related observations that also document these points have been reported by Stein (1978), and this transfer-appropriate processing approach has been extended to differences between explicit and implicit memory tests (Blaxton, 1989; Roediger, 1990; see Kelley & Lindsay, this volume, Chapter 2).

Glisky and Rabinowitz (1985) have also shown that repetition of mental operations at test enhances memory performance (recognition in their case). In their study phase, subjects either read intact words (alcohol) or generated them from word fragments in which two letters were omitted (al—oho—). The words were easy to generate with letters missing so subjects made relatively few generation errors. When subjects took a later test, they either saw intact words or generated them from the same fragment, just as they had at study. The recognition test required subjects to decide if the words they were reading and generating on the test had appeared in the original study list (in either form). Results are presented in Table 9. The left column shows the standard generation effect (Jacoby, 1978; Slamecka & Graf, 1978): Generated words were recognized better than read words when the recognition test involved reading intact words. However, examination of the data in the right column reveals that the generation effect was enhanced by about 10% when the generation operations were reinstated at the time of test. Best recognition occurred when subjects generated words at study and test. Note that generating at test did not have a general advantage, but only enhanced recognition when subjects had generated the words during study. In companion experiments, Glisky and Rabinowitz showed that it was necessary for the same operations to be carried out at study and test (i.e., the same

TABLE 9 Results of Glisky and Rabinowitz (1985)^a

Encoding condition	Retrieval condition	
	Read	Generate
Read	.60	.59
Generate	.76	.86
Difference (G - R)	.16	.27

^aSubjects either read or generated single words at study and then during a later recognition test. False alarm rates were the same for Read (.18) and Generated (.19) words. Results shown are the proportion of hits.

letters to be omitted from the words) to see the whole advantage of generating items at test. In general, Glisky and Rabinowitz “found that the more closely the operations at retrieval matched those at encoding, the better [was] recognition performance” (p. 193).

A striking demonstration that corroborates the finding of Glisky and Rabinowitz (1985) has recently been reported by Engelkamp, Zimmer, Mohr, and Sellen (1994). They compared subjects’ recognition of phrases that were either verbally encoded or self-performed. That is, subjects were given brief statements such as “close the book” and “pick up the pencil,” and they either simply read the phrase or they actually performed the task with materials in front of them. Previous research had shown that subjects remembered the events better if they performed them than if they merely read about them. Engelkamp et al. wondered if this effect might be magnified if subjects were asked to perform the task again at test. (The standard test that had been used in prior research involved recognition of the verbal form of the item.) Therefore, during the study phase, they had subjects either read statements, or read them and perform the actions. On a later recognition test, subjects either read the statements, or read and performed them, before judging if they had been previously studied. Some test items referred to previously studied items and some were new, to measure the false alarm rate.

The results are presented in Table 10. The verbal test condition (reading the statement) showed a sizable advantage for self-performed tasks at study, 23% in Experiment 1 and 26% in Experiment 2, relative to verbal tasks (reading the statement at study), reflecting the typical advantage accruing to self-performed tasks. However, in both experiments, the advantage of self-performed tasks was increased if the tasks were also self-performed during the test, increasing the size of the effect by 7% in Experiment 1 and 12% in Experiment 2. Once again, recapitulating operations between study and test enhanced performance, although in this case the operations were actual physical movements and not mental operations. In another experiment,

TABLE 10 Results of Engelkamp, Zimmer, Mohr, & Sellen (1994)^a

Encoding condition	Retrieval condition			
	Experiment 1		Experiment 2	
	Verbal	Self-performed	Verbal	Self-performed
Verbal	.64 ^b	.64	.45	.40
Self-performed	.87	.94	.71	.78
Difference (SP - V)	.23	.30	.26	.38

^aSubjects read phrases (verbal encoding) or in addition performed the action suggested by the phrase (self-performed encoding). The recognition test was composed of both old and new phrases that subjects read or performed.

^bData are corrected recognition (hits minus false alarms).

Engelkamp et al. (1994) showed that enhanced recognition of self-performed tasks was diminished if subjects performed the task with one hand at study and then with a different hand at test (e.g., opening the book with the right hand at study and then opening it with the left hand at test). Therefore, the motor component of recognition seems rather specific.

In general, the results reviewed in this section show quite clearly that the means of acquisition of material forms part of its representation in memory and that to the extent that the retrieval task recapitulates this original enactment, performance benefits. All these results are clearly in line with Kolers's (1979; Kolers & Roediger, 1984) procedural account of recognition performance.

F. Relational and Item-Specific Processing

At several points in the chapter we have written about the importance of distinctiveness, because, in general, distinctive events are well remembered. This effect has been known at least since the time of von Restorff (1933) and has been studied in many different paradigms since then (Guynn & Roediger, 1995; Tulving, 1969; Wallace, 1965; see Hunt & McDaniel, 1993, for review). The basic idea, embedded in many different theories, is that at least two different ways of encoding information exist (J. R. Anderson, 1972; Hunt & Einstein, 1981; Mandler, 1980). In one case, people can pay attention to the item-specific aspects of a particular event during encoding. This increases the likelihood that memory for that event will include features that differentiate the event from others in memory, thus making the event distinctive. The other type of processing is relational, which serves to interrelate events or to organize them according to some scheme. One basic idea is that item-specific processing will primarily enhance recognition, whereas both types of processing will benefit recall. In free recall, the inter-

relations or associations among events help to cue subjects so that recall of some events will cause people to remember more events later.

The basic importance of this distinction was illustrated in an important study by Hunt and Einstein (1981, Experiment 1). Subjects studied either a categorized list of 36 words (6 words from each of 6 categories), or a list of 36 words that subjects perceived to be unrelated (6 words from each of 6 ad hoc categories, e.g., liquids, things that are green, things that fly). Hunt and Einstein assumed that *related* words would automatically receive relational processing, but not necessarily item-specific processing; further, words that subjects perceived to be unrelated were assumed to automatically receive item-specific processing, but not necessarily relational processing. Subjects performed either a relational orienting task (sorting the items into specified categories) or an item-specific orienting task (rating the pleasantness of the items) on either the related or the unrelated words. Sorting the words into categories should require relational processing (for both the related and unrelated words), and rating the words for pleasantness should require item-specific processing (for both the related and unrelated words). The idea was that the most beneficial type of processing might be that which was *different from the processing invited by the list structure*; that is, the sorting task might primarily benefit memory for the unrelated words, because for the unrelated words, information gained from rating items for pleasantness is redundant with information from the list structure (i.e., both encourage item-specific information). Conversely, the pleasantness rating task might primarily benefit memory for the related words, because, for the related words, information gained from sorting is redundant with information from the list structure (i.e., both encourage relational information). Of course, the effects of these variables might well depend on the type of memory test. Thus, after either sorting or rating a *related* or an *unrelated* list, subjects were given tests of free recall (thought to depend on both relational and item-specific information) and recognition (thought to depend mainly on item-specific information).

The results appear in Table 11. First, consider recall results, shown on the left side of the table. As predicted, the item-specific task produced better free recall than did the relational task for the related list (.48 vs. .42 of the words recalled), but the relational task produced better free recall than did the item-specific task for the unrelated list (.47 vs. .33 of the words recalled). For the related list, information gained from rating (i.e., item-specific information) was not redundant with the information gained from the interrelations among list members, but relational information gained from sorting was somewhat redundant and thus less useful. For the unrelated list, information gained from sorting (i.e., relational information) was not redundant with the information gained from the list of unrelated words, but item-specific information gained from rating was perhaps redundant

TABLE 11 Results of Hunt and Einstein's Experiment 1 (1981)^a

Encoding task	Retrieval condition			
	Free recall		Recognition ^b	
	List type		List type	
	Related	Unrelated	Related	Unrelated
Relational (sorting)	.42	.47	.73	.89
Item-specific (rating)	.48	.33	.93	.91

^aSubjects studied 36 related or nominally unrelated words, sorted the words into categories (relational task) or rated the pleasantness of each word (item-specific task), and took tests of free recall and recognition.

^bRecognition scores are AG scores. The AG score is a nonparametric signal detection measure (Pollack, Norman, & Galanter, 1964).

and thus less useful. These results show that both item-specific and relational processing are necessary for free recall.

The recognition results (on the right side of Table 11) differed somewhat from the recall results, however. For the related list, the item-specific task did produce better recognition than did the relational task (.93 vs. .73), the same pattern as in free recall. But for the unrelated list, the item-specific and relational tasks produced similar recognition (.91 vs. .89), with a slight benefit for the item-specific task, even though the words were nominally unrelated. However, this equivalence in performance should be treated cautiously, as performance is near the ceiling.

Hunt and Einstein's (1981) results provide another example of an encoding/retrieval interaction: The effect of a study variable (sorting or pleasantness rating) depended on the type of list (related or unrelated) as well as on the type of test (free recall or recognition). Sorting produced better free recall for the nominally unrelated list, and pleasantness rating produced better free recall for the related list. Rating the pleasantness of words produced better recognition for the related list, but for an unrelated list, rating and sorting produced similar recognition scores.

These ideas—relational and item-specific processing—are useful in explaining a number of phenomena in the literature. Relational processing can provide an account for organizational bases of recall (J. R. Anderson, 1972; Mandler, 1967; Tulving, 1968). For present purposes, item-specific or distinctive processing can be understood in terms of cue overload; that is, a distinctive event, such as a picture embedded in a list of words, is well remembered because subjects trying to retrieve items from the category "pictures" have only one item to retrieve. The same mechanism may be at work to produce what are called flashbulb memories (R. Brown & Kulik,

1977), which are believed to be extraordinarily vivid memories associated with a particular cataclysmic event. Most of us remember vividly (or at least *believe* we remember vividly) events surrounding a particularly emotional event, whether personal (the birth of a child), or a national tragedy, such as the explosion of the space shuttle Challenger or the assassination of a president. These distinctive events are unique and, therefore, serve as powerful retrieval cues, with no other memories associated to the event. We can thus understand how these memories continue to stand out and be retrievable even when more recent memories have suffered interference from overloaded cues.

G. Reprise

This chapter has emphasized principles of transfer-appropriate processing (or encoding specificity) and distinctiveness (or absence of cue overload). We culminate this section with an apt experiment showing how good retrieval can be if a situation maximizes both principles. Mäntylä (1986) reported several experiments in which he presented subjects with 500–600 words, under incidental learning conditions, with subjects showing 90% retention on a later test of recall! How did his subjects achieve this level of performance on such a huge amount of material? In Mäntylä's experiments, the subject's encoding task was to generate either one property or three properties that described or defined each word. So, for example, when given the word *marmalade*, subjects might write that it is *sticky*, *sweet*, and *can be eaten*. Of course, it took subjects considerable time to generate these properties. During the later test, subjects were given either a single cue or all three cues that they had generated. Mäntylä showed that subjects given a single cue could remember about 50% to 60% of the words, and those given three cues could remember about 90% on a test given the same day as the study phase. This level of performance was not achieved by guessing because an appropriate control group, who did not receive the encoding phase but who were given cues and told to generate the words, performed at the level of 5% with one cue and 17% with three cues. Guessing may have played some role in the subjects' spectacular performance, but the greater part seems due to other factors powerfully aiding their memories.

Mäntylä (1986) argued for two factors. One is cue/trace compatibility, which embodies the logic of the encoding specificity hypothesis or transfer-appropriate processing. The retrieval cues were highly effective because they were quite compatible with the subject's mode of encoding the material: The subjects themselves had constructed the cues during encoding. The second principle was cue distinctiveness: The cues subjects generated were quite different from one another, each tailored to the particular word-event to be remembered. Because the cues precisely identified the targets, the

item-specific cues had little or no overlap with other cues, and were thus distinctive. Therefore, Mäntylä's study serves as a powerful demonstration of the importance of the two principles guiding this chapter: transfer-appropriate processing (or encoding specificity) and distinctive, item-specific processing. If factors in a situation maximize the effectiveness of both these principles, retention will be outstanding, as it was for Mäntylä's subjects.

V. EFFECTS OF PRIOR RETRIEVAL

Tests of memory are not neutral events that merely assess the state of a person's knowledge; tests also change or modify memories (Bjork, 1975; McDaniel & Masson, 1985; see also Dempster, this volume, Chapter 9, for a discussion of test effects). Changes can be either positive, aiding later retention, or negative, causing forgetting, interference, and even false recollections. We consider some positive effects here and negative effects in a later section of the chapter.

A. Effects of One Test on Another

How does one test of memory affect performance on another test given later? Does a recall test improve later recognition? Does a recognition test improve later recall? The answer to this last question is relatively easy, because the recognition test also serves as another study opportunity: The *old* words on a recognition test are presented again, and thereby benefit subsequent recall of those words. Although the evidence is not entirely consistent, a recall test also generally improves subsequent recognition, at least for the last few items in a list (Jones & Roediger, 1995; Lockhart, 1975). Other researchers have asked more subtle questions with regard to the influence of one test on another, and the results are complex (McDaniel, Kowitz, & Dunay, 1989). These questions reflect the fact that tests differ on dimensions other than whether they require recall or recognition. For instance, as illustrated in the Morris et al. (1977) study described earlier (as well as in McDaniel et al., 1978), tests differ on the kind of information required for successful performance. For example, depending on how it is constructed, accurate responding on a recognition test might require access to meaning, to phonemic features (Morris et al., 1977), or to graphemic (visual appearance of letters) features (McDaniel et al., 1978). In line with the notion of levels of processing, prior tests that require use of semantic information tend to benefit later tests more than do prior tests that require use of nonsemantic (e.g., phonemic) information (Bartlett, 1977; McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985).

The pattern becomes more complex, although understandable from the retrieval principles already presented, when one considers the nature of the

information tested on the later test. McDaniel, Kowitz, and Dunay (1989, Experiment 2) varied whether the initial test focused on semantic or phonemic information by providing subjects with meaningful associates or rhymes, respectively, of studied words (i.e., a cued recall test was used). On a later test, subjects were provided with either identical cues from the earlier test, different cues directed at the same level of information as the cues from the earlier test, or different cues directed at a different level of information from the cues used on the earlier test. The initial test most improved performance on the later test (relative to conditions in which the words were not tested initially) to the degree that the later test focused on information that matched the information used in the earlier test. That is, identical cues were better than different cues directed at the same level of information, which in turn were better than different cues directed at a different level of information. This pattern reflects a transfer-appropriate processing effect, and suggests that memory testing induces additional processing of already-encoded information. As a general conclusion, then, there is a benefit from taking a prior test on taking a later test of whatever sort, although the amount of benefit will differ depending on the nature of the two tests (McDaniel, Kowitz, & Dunay, 1989).

B. The Testing Effect: Taking the Same Test Twice

From the perspective of transfer-appropriate processing, and the results just discussed, we might expect that the greatest facilitation on a later test would occur if subjects had taken the same test previously and successfully retrieved the events. What could be better practice for retrieving information on a later test than having retrieved it earlier? Indeed, often a prior test has a greatly facilitating effect on a later test, even better than another study opportunity for the material. This finding has a long history in experimental psychology (Spitzer, 1939), but its importance for educational settings seems overlooked (but see McDaniel & Fisher, 1991). The testing effect, as it is called, occurs both for meaningful prose materials and for lists of words in laboratory settings (Glover, 1989). To provide but one example, Wheeler and Roediger (1992) had subjects study a series of 60 pictures under one of two conditions, either in the context of a story presented auditorily (the pictures popped up as the appropriate words occurred in the story), or simply in a list where the names of the pictures were heard as the pictures were presented. After the pictures were presented in one of the two encoding conditions, subjects filled out a brief questionnaire and then took either zero, one, or three successive free-recall tests on the studied pictures (subjects wrote down the names of the pictures on the test). For present purposes, the critical test of interest occurred when subjects returned to the laboratory a week later and took one last free-recall test.

The results of this last test are shown in Figure 2, where it can be seen that the pictures embedded in the story were generally remembered better than pictures presented only with their names. That is, the story did promote better recall. However, this later benefit required an earlier test to bring it out. The number of prior tests given a week before greatly affected memorability on the delayed test, with the more powerful effect occurring for pictures that were recalled in the context of the story rather than presented in a list. Because the three groups of subjects in each of the two encoding conditions (pictures plus story or pictures plus names) had been treated identically up to the point of the first test, the large differences in retention a week later must have been a direct consequence of the number of tests taken. In general, the retrieval practice induced by taking a prior test provided great transfer or benefit to taking a later test. Indeed, many experiments have shown that retrieval practice from taking a test actually produces greater gains in later retention than does another presentation of the material for study (Hogan & Kintsch, 1971; McDaniel & Masson, 1985; Thompson, Wenger, & Bartling, 1978; Tulving, 1967). Runquist (1986) showed that a test can actually eliminate forgetting in some conditions.

A final example of the power of testing comes from a study by Landauer and Bjork (1978), who examined different schedules of testing in recall of

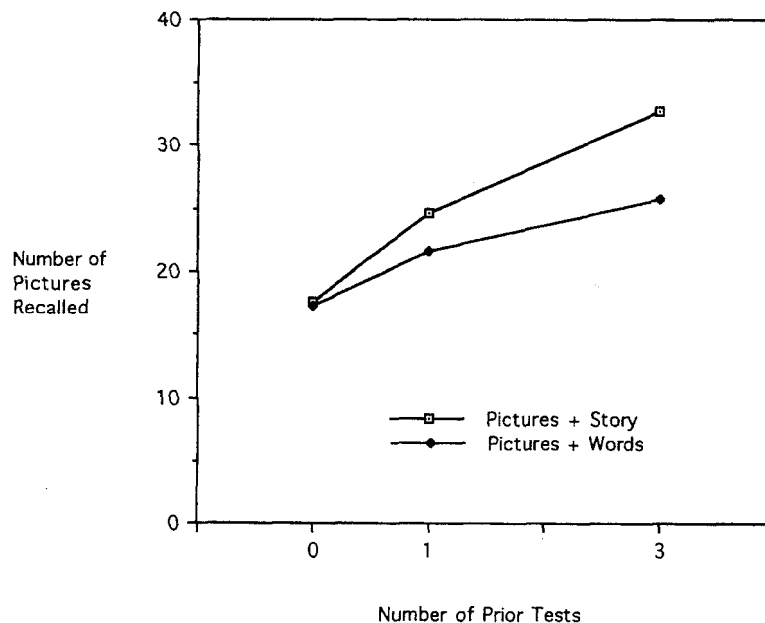


FIGURE 2 The testing effect. The number of tests taken soon after studying pictures greatly affected recall one week later. (Data from Wheeler and Roediger, 1992.)

name–face pairs. Is it better to have a test (with feedback) shortly after a pair has been presented, or is it better to wait? Landauer and Bjork tested several variations and discovered that a schedule of temporally expanding tests produced the best performance on a delayed retention test. That is, in remembering name–face pairs, subjects benefited most when they were first tested after a short delay (before they had forgotten the pair), then after a bit longer delay, then after a still longer delay, and so on. Because the eventual test of retention occurs at a great delay (hours, days, or weeks later), this expanding test procedure helps to shape eventual retention under transfer-appropriate conditions: The best practice for a long delayed retention test is being able to retrieve the pair after increasingly long delays.

VI. RELATED TOPICS

In a brief chapter on retrieval processes, we cannot hope to cover in full all the topics falling under its purview. Yet we would be remiss if we did not deal, however superficially, with some other miscellaneous topics.

A. Repression

Early in this century, Freud (1914/1957) popularized the idea that some memories are too painful and traumatic to be permitted into consciousness. Such memories are banished to an unconscious state, where they may nevertheless disrupt ongoing behavior in the form of neurotic symptoms, slips of the tongue, or expression in dreams. The concept of repressed memories existing in the unconscious is firmly embedded in the folk psychology of our times, but experimental evidence for it is slight. Indeed, we know of no strong evidence for the existence of unconscious memories from repression, at least in the strong sense. If, instead, repression is defined simply as people's trying to avoid thinking about unpleasant topics (Erdelyi & Goldberg, 1979), then certainly it exists. On the other hand, the notion of traumatic memories being banished to an unconscious state and then suddenly reappearing is supported only by anecdotes that cannot be experimentally evaluated. In fairness, however, experimental tests of the repression idea are notoriously difficult to undertake. Nonetheless, those who believe in repression and repressed memories do so as an article of faith rather than as a tested proposition having an experimental basis in fact (Loftus, 1993).

An idea akin to that of repression is that of retrieval inhibition (Bjork, 1989). One idea is that people can inhibit material that is no longer necessary—they can tell themselves to forget it or be so instructed by an experimenter—and this material will then not create interference with other material. Some researchers have specifically linked retrieval inhibition as studied in the laboratory with other forms of inhibited or repressed memories

(Geiselman, Bjork, & Fishman, 1983). Such inhibitory processes are considered more fully by Anderson and Neely in Chapter 8 of this volume. However, the link between retrieval inhibition as studied in the laboratory and the Freudian concept of repression remains tenuous at best.

B. Hypnosis and Memory Retrieval

A companion idea to that of repression is the notion that there exist psychological means of unlocking repressed memories and bringing them to the surface. Hypnosis is considered by some practitioners to be one of these means. Hypnotized subjects are placed in a relaxed state and encouraged to visualize the original event. On some occasions, hypnotized subjects do recall information that they could not previously recall in an awake state, but at the expense of increased guesses (Dywan & Bowers, 1983). Many laboratory experiments have been carried out, both with word and picture lists and with materials that more realistically approximate natural events, but with the same outcome: There is no good evidence that hypnosis can aid memory retrieval (M. C. Smith, 1983). Gains that are seen under hypnosis in some experiments would likely occur merely from the act of repeated testing (Nogrody, McConkey, & Perry, 1985). In other cases where hypnosis appears successful, the outcome is likely due to the instructions and cues given during hypnosis on how to retrieve information, rather than on hypnosis per se (Geiselman, Fisher, MacKinnon, & Holland, 1985). Although some psychological techniques may enhance retrieval from memory, hypnosis does not seem to have this power.

C. Spontaneous Recovery

In classical conditioning experiments, a conditioned stimulus (CS) can be made to elicit a conditioned response (CR) through repeated pairings with an unconditioned stimulus (US). After conditioning, if the conditioned stimulus is repeatedly presented in the absence of the unconditioned stimulus, the conditioned response declines, and eventually disappears altogether (extinction). However, if the researcher waits for a period of time and then re-presents the conditioned stimulus, the conditioned response recovers. Such recovery occurs even in the absence of any additional CS-US pairings and, therefore, is called spontaneous recovery. Pavlov (1927) first discovered this phenomenon, and researchers in human learning and memory have sought analogues ever since.

In humans, the question becomes: *Can memory improve over time in the absence of repeated testing?* (We have already seen—through the phenomena of reminiscence and hypermnesia—that subjects will often be able to remember more on a later test than on an earlier test; the question with

regard to spontaneous recovery is whether the increase can occur over time without intervening testing.)

The evidence on spontaneous recovery is mixed, but it does seem to occur. The necessary condition in humans is to have a large amount of interference that blocks retrieval of events. In a typical paradigm, subjects learn paired associates (A–B pairs, such as *radio–hammer*) and then later they learn other pairs (A–D, *radio–dictionary*) with the same stimuli and different responses. After the interpolated A–D learning, subjects are then given the stimulus (A, or *radio*) and asked for the response from the first list. If the test is given immediately after learning the interpolated list, subjects show great amounts of retroactive interference, and recall of B (*hammer*) is poor. If an interval elapses between the interpolated learning and the test of the first list association, however, then recall of B (*hammer*) is better (at least in some experiments). This improvement is also referred to as spontaneous recovery and seems at least analogous to the process occurring in Pavlovian conditioning. Barnes and Underwood (1959) presented evidence for such spontaneous recovery, although other researchers have had difficulty replicating the effect. In a thorough review of this literature, however, A. S. Brown (1976) concluded that the bulk of the evidence supported spontaneous recovery. In addition, Wheeler (1995) has also reliably obtained spontaneous recovery in several paradigms in a recent series of experiments. Thus, the phenomenon does seem to exist, but it remains poorly understood.

D. Retrieval of False Memories

Other chapters in this volume are concerned with effects of misleading information on memory, but here we pause to note that the act of retrieval can itself cause forgetting and false memories in several ways. The phenomenon of output interference (Tulving & Arbuckle, 1963) indicates that the process of recalling some events may interfere with recall of later events. This process has been most thoroughly studied in categorized list recall, where recall of early categories seems to inhibit recall of later categories (Roediger, 1978; see also M. C. Anderson, Bjork, & Bjork, 1994) and recall of the contents of those categories (Roediger & Schmidt, 1980).

Retrieval of events can modify memory in other ways, too. If subjects retrieve an event from memory—even if they are in error and know it at the time—then often they will falsely recollect this event as a “true” memory later on. Roediger, Wheeler, and Rajaram (1993) reported an experiment in which subjects were forced to guess items on a first test and then, unexpectedly, were tested again a week later. They discovered that subjects often falsely recalled their guesses from the first test as items actually appearing in the study list. This sort of error indicates a reality monitoring problem of the type studied by Johnson and Raye (1981). The general point of this

section is that the act of retrieval does not always enhance performance (as in the testing effect discussed earlier), but may harm memory through output interference and in other ways. These harmful effects are discussed further by Anderson and Neely and by Chandler and Fisher in Chapters 8 and 14, respectively.

VII. CONCLUSION

Tulving (1974) emphasized that remembering is a product of information from two sources: encoded information or “memory traces” and retrieval information. *Product* is to be taken literally in this treatment: Relatively weak traces may be accessed by powerful retrieval cues, but with strong traces, weak cues may suffice. The study of retrieval processes cannot occur in isolation any more than can the study of memory traces. Memory traces must be actualized through retrieval, and retrieval without memory traces is confabulation.

We have argued in this chapter that consideration of two principles helps us, at our current primitive level of understanding, to organize and systematize knowledge of memory retrieval. One principle, stated in various ways by different investigators, is that performance on a retention test will benefit to the extent that the operations required by the test match those used in initial encoding or learning of the event (the encoding specificity principle, or the principle of transfer-appropriate processing). The second principle is that a cue will be effective to the extent that it is distinctive or not overloaded by numerous memories (the cue overload principle).

Acknowledgments

Preparation of this chapter was supported by Grant AF 49620-92-J-0437 from the Air Force Office of Scientific Research to Henry L. Roediger, III, and by a National Science Foundation Graduate Research Fellowship to Melissa J. Guynn. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

We thank Mark A. McDaniel for his aid in preparing the chapter and Elizabeth Bjork, Robert Bjork, Eric Eich, Dawn Macaulay, and Endel Tulving for helpful comments.

References

- Anderson, J. R. (1972). FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 315–378). New York: Academic Press.
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Exploring the retrieval dynamics of long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.

- Ballard, P. B. (1913). Oblivescence and reminiscence. *British Journal of Psychology Monograph Supplements*, 1, 1–82.
- Barclay, J. R., Bransford, J. D., Franks, J. J., McCarrell, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, 13, 471–482.
- Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first list associations in transfer theory. *Journal of Experimental Psychology*, 58, 97–105.
- Bartlett, J. C. (1977). Effects of immediate testing on delayed retrieval: Search and recovery operations with four types of cue. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 719–732.
- Bilodeau, E. A., & Blick, K. A. (1965). Courses of misrecall over long-term retention intervals as related to strength of preexperimental habits of word association. *Psychological Reports*, 16, 1173–1192.
- Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.
- Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.
- Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane-Flowerree symposium on cognition* (pp. 313–344). Hillsdale, NJ: Erlbaum.
- Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin*, 99, 229–246.
- Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 657–668.
- Bower, G. H. (1981). Mood and memory. *American Psychologist*, 36, 129–148.
- Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 331–354). Hillsdale, NJ: Erlbaum.
- Brown, A. S. (1976). Spontaneous recovery in human learning. *Psychological Bulletin*, 83, 321–338.
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, 5, 73–99.
- Brown, W. (1923). To what extent is memory measured by a single recall trial? *Journal of Experimental Psychology*, 6, 377–382.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268–294.
- Dywan, J., & Bowers, K. (1983). The use of hypnosis to enhance recall. *Science*, 222, 184–185.
- Earhard, M. (1967). Cued recall and free recall as a function of the number of items per cue. *Journal of Verbal Learning and Verbal Behavior*, 6, 257–263.
- Eich, J. E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition*, 8, 157–173.
- Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 764–770.
- Eich, E. (1989). Theoretical issues in state-dependent memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 331–354). Hillsdale, NJ: Erlbaum.
- Eich, E., Macaulay, D., & Ryan, L. (1994). Mood dependent memory for events of the personal past. *Journal of Experimental Psychology: General*, 123, 201–215.

- Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 443–455.
- Eich, J. E., Weingartner, H., Stillman, R. C., & Gillin, J. C. (1975). State dependent accessibility of retrieval cues in the retention of a categorized list. *Journal of Verbal Learning and Verbal Behavior*, 14, 408–417.
- Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition*, 22, 34–39.
- Erdelyi, M. H. (1984). The recovery of unconscious (inaccessible) memories: Laboratory studies of hypermnesia. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 95–127). New York: Academic Press.
- Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not for words in multiple recall trials. *Cognitive Psychology*, 6, 159–171.
- Erdelyi, M. H., & Goldberg, B. (1979). Let's not sweep repression under the rug: Toward a cognitive psychology of repression. In J. F. Kihlstrom & F. J. Evans (Eds.), *Functional disorders of memory* (pp. 355–402). Hillsdale, NJ: Erlbaum.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, 62, 145–154.
- Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition*, 13, 333–345.
- Fisher, R. P., & Craik, F. I. M. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 701–711.
- Freud, S. (1957). The history of the psychoanalytic movement. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 14). London: Hogarth Press. (Originally published 1914)
- Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, 112, 58–72.
- Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, 70, 401–412.
- Glisky, E. L., & Rabinowitz, J. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 193–205.
- Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392–399.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66, 325–331.
- Guynn, M. J., & Roediger, H. L. (1995). The impact of distinctive events on implicit and explicit tests of memory. *Psychological Research*, 57, 192–202.
- Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning*. Englewood Cliffs, NJ: Prentice-Hall.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 562–567.
- Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, 20, 497–514.
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, 32, 421–445.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, 17, 649–667.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.

- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.
- Jones, T. C., & Roediger, H. L. (1995). The experiential basis of serial position effects. *European Journal of Cognitive Psychology*, *7*, 65–80.
- Köhler, W. (1947). *Gestalt psychology*. New York: Liverwright.
- Kolers, P. A. (1979). A pattern-analyzing basis of recognition. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 363–384). Hillsdale, NJ: Erlbaum.
- Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, *23*, 425–449.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.
- Lockhart, R. S. (1975). The facilitation of recognition by recall. *Journal of Verbal Learning and Verbal Behavior*, *14*, 253–258.
- Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist*, *48*, 518–537.
- Loftus, E. F., & Loftus, G. R. (1980). On the permanence of information stored in the human brain. *American Psychologist*, *35*, 409–420.
- Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 327–372). New York: Academic Press.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271.
- Mäntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 66–71.
- McDaniel, M. A., Anderson, D. C., Einstein, G. O., & O'Halloran, C. M. (1989). Modulation of environmental reinstatement effects through encoding strategies. *American Journal of Psychology*, *102*, 523–548.
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*, 192–201.
- McDaniel, M. A., Friedman, A., & Bourne, L. E. (1978). Remembering the levels of information in words. *Memory & Cognition*, *6*, 156–164.
- McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition*, *17*, 423–434.
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371–385.
- McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, *39*, 352–370.
- Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*, 1–21.
- Mensink, G. J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434–455.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.
- Murdock, B. B. (1989). The past, the present, and the future: Comments on section 1. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 93–98). Hillsdale, NJ: Erlbaum.
- Newman, S. E., & Frith, U. (1977). Encoding specificity vs. associative continuity. *Bulletin of the Psychonomic Society*, *10*, 73–75.
- Nograd, H., McConkey, K. M., & Perry, C. (1985). Enhancing visual memory: Trying hypnosis, trying imagination, and trying again. *Journal of Abnormal Psychology*, *94*, 195–204.
- Pan, S. (1926). The influence of context upon learning and recall. *Journal of Experimental Psychology*, *9*, 468–491.

- Pavlov, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). London: Oxford University Press.
- Payne, D. G. (1987). Hypermnnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, *101*, 5–27.
- Pollack, I., Norman, D. A., & Galanter, E. (1964). An efficient nonparametric analysis of recognition memory. *Psychonomic Science*, *1*, 327–328.
- Roediger, H. L. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, *12*, 644–657.
- Roediger, H. L. (1978). Recall as a self-limiting process. *Memory & Cognition*, *6*, 54–63.
- Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*, 1043–1056.
- Roediger, H. L., & Adelson, B. (1980). Semantic specificity in cued recall. *Memory & Cognition*, *8*, 65–74.
- Roediger, H. L., & Challis, B. H. (1989). Hypermnnesia: Improvements in recall with repeated testing. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane-Flower symposium on cognition* (pp. 175–199). Hillsdale, NJ: Erlbaum.
- Roediger, H. L., & Payne, D. G. (1983). Superiority of free recall to cued recall with “strong” cues. *Psychological Research*, *45*, 275–286.
- Roediger, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired associate lists. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 91–105.
- Roediger, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnnesia. *Memory & Cognition*, *6*, 296–305.
- Roediger, H. L., Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing and reconstructing the past. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 30, pp. 97–134). New York: Academic Press.
- Runquist, W. N. (1986). The effect of testing on the forgetting of related and unrelated associates. *Canadian Journal of Psychology*, *40*, 65–76.
- Saufley, W. H., Otaka, S. R., & Bavaresco, J. L. (1985). Context effects: Classroom tests and context independence. *Memory & Cognition*, *13*, 522–528.
- Schab, F. R. (1990). Odors and the remembrance of things past. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 648–655.
- Schacter, D. L. (1982). *Stranger behind the engram*. Hillsdale, NJ: Erlbaum.
- Schacter, D. L., Eich, J. E., & Tulving, E. (1978). Richard Semon’s theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *17*, 721–744.
- Semon, R. (1921). *The mneme*. London: Allen & Unwin.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.
- Smith, M. C. (1983). Hypnotic memory enhancement of eyewitnesses: Does it work? *Psychological Bulletin*, *94*, 387–407.
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 460–471.
- Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13–34). Chichester, England: Wiley.
- Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342–353.
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656.
- Stein, B. S. (1978). Depth of processing reexamined: The effects of precision of encoding and test appropriateness. *Journal of Verbal Learning and Verbal Behavior*, *17*, 165–174.

- Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 210–221.
- Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 86, 255–262.
- Tulving, E. (1967). The effects of presentation and recall in free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 6, 175–184.
- Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2–36). Englewood Cliffs, NJ: Prentice-Hall.
- Tulving, E. (1969). Retrograde amnesia in free recall. *Science*, 164, 88–90.
- Tulving, E. (1974). Cue-dependent forgetting. *American Scientist*, 62, 74–82.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- Tulving, E. (1991). Interview. *Journal of Cognitive Neuroscience*, 3, 89–94.
- Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1, 321–334.
- Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 77, 593–601.
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5, 381–391.
- Tulving, E., & Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352–373.
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologie Forschung*, 18, 299–342.
- Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, 63, 410–424.
- Watkins, M. J. (1979). Engrams as cuegrams and forgetting as cue overload: A cueing approach to the structure of memory. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 347–372). New York: Academic Press.
- Watkins, M. J., & Gardiner, J. M. (1982). Cued recall. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 173–195). New York: Academic Press.
- Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, 104, 442–452.
- Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72, 89–104.
- Weiner, B. (1966). Effects of motivation on the availability and retrieval of memory traces. *Psychological Bulletin*, 65, 24–37.
- Wheeler, M. A. (1995). Improvement in recall without repeated testing: Spontaneous recovery revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 173–184.
- Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240–245.
- Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, 1, 89–106.