Imagery and Related Mnemonic Processes
Theories, Individual Differences, and Applications

Edited by
Mark A. McDaniel
Michael Pressley

With 23 Illustrations

Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo
CHAPTER 7

Reversing the Picture Superiority Effect

Henry L. Roediger III and Mary Susan Weldon

Imaginal coding typically enhances retention. Pictures are remembered better than words; words for which subjects imagine referents are better remembered than words studied without such coding; concrete words are better retained than abstract words; and mnemonic devices employing imagery can produce dramatic effects on retention. These facts have long been noted (Paivio, 1971) and many contributions in this volume confirm the efficacy of imagery and imagination in remembering.

If we conceive the process of remembering as involving the broad stages of encoding, storage, and retrieval (Köhler, 1947; Melton, 1963), then it is fair to say that theories explaining the superior retention of information encoded imaginally have emphasized encoding and storage processes as the loci of the effects. For example, Paivio’s (1969, 1986) dual code theory attributes the picture superiority effect to redundant coding, and Nelson’s (1979) sensory-semantic model to pictures’ superior sensory codes. Interestingly, the possible importance of retrieval factors has been relatively neglected.

In this chapter we present evidence that the neglect of the retrieval phase in such effects is unwarranted. In fact, we show that retrieval factors are critical in producing the “standard” imagery effects. We concentrate primarily on the picture superiority effect, the finding that pictures are better remembered than words, but at the end of the chapter we argue that our thesis also holds for other standard effects of imagery and mnemonic devices.

First, we briefly review evidence documenting the picture superiority effect. Next, in something of a detour to lay the groundwork of our argument, we review recent evidence showing important differences between certain classes of retention tasks. From this literature we hypothesize that one may in fact find a reversal of the usual picture superiority effect on certain tests, such that words are retained better than pictures. We present several experiments that illustrate superior retention for words than for pictures and then resurrect some older evidence bearing on this point. We conclude with a section summarizing our argument and extending it to other imagery manipulations and mnemonic techniques.
The Picture Superiority Effect

The first evidence suggesting that pictures or objects are remembered better than words was produced in a study by Kirkpatrick in 1894. He presented subjects with either objects or words and tested retention both immediately and 72 h later. At both intervals he found that the objects were better remembered than the words. This finding was replicated by Calkins (1898), who compared retention of pictures and words under better controlled conditions, but was then largely ignored by researchers interested in human learning and memory for some 60 years, until the important line of work initiated by Allan Paivio in the 1960s. In the past 20 years the superior retention of pictures relative to words on many tests of episodic memory has been reported with heartening regularity for a science often plagued with failures to replicate. Here we touch on only some features of this evidence and refer the reader to excellent discussions by Madigan (1983) and Paivio (1971, Chapter 7; 1986, Chapter 8) for more thorough reviews.

The general conclusion to emerge from the literature is that when subjects study a list of pictures and words, pictures are better remembered on tests of free recall (e.g., Paivio, Rogers, & Smythe, 1968), recognition (e.g., Madigan, 1983), serial recall (Herman, Broussard, & Todd, 1951), and paired-associate learning (Paivio & Yarmey, 1966), although some qualifications exist for each of these tasks. The general form of the qualification is that the picture superiority effect diminishes or evaporates under certain encoding conditions. Thus, for example, elaborative semantic or imaginal encoding of words can produce word recall equal to or better than picture recall (Durso & Johnson, 1980). At a very fast rate of presentation (5.3 items/sec), free recall of pictures and words is equivalent and serial recall of words exceeds that of pictures (Paivio & Csapo, 1969). In paired-associate learning pictures (P) produce better performance than words (W) when used as stimuli (P-P and P-W pairs lead to better performance than W-P and W-W pairs) but, when used as responses, pictures show no advantage to words and sometimes even produce slightly worse performance (see Paivio, 1971, p. 255; Postman, 1978). These exceptions must be viewed against a background of studies typically producing robust picture superiority effects. A typical example in free recall, from Paivio et al. (1968), is shown in the left panel of Figure 7.1.

Theorists' emphasis on encoding operations as the locus of the picture superiority effect seems natural, because in previous research with standard measures, variation in form of the test alone has never eliminated the picture superiority effect. That is, we can find no experiments using customary memory tests in which study conditions have been held constant and various forms of test have yielded both the picture superiority effect and its elimination or reversal across tests.

Recognition memory experiments with words as the test items are the most informative on this issue, because the format of the test item matches the study episode more closely for words than for pictures. Many popularly accepted notions about retrieval processes, such as the encoding specificity principle (Tulving & Thomson, 1973), would seem to predict that on a word recognition test, words should be remembered better than pictures. This is because there is
7. Reversing the Picture Superiority Effect

![Figure 7.1](image)

**Figure 7.1.** The picture superiority effect as shown in free recall (left panel; results from Paivio et al., 1968) and in word recognition (right panel; results from Madigan, 1983).

a greater match between the study and test episodes in the (study) word−(test) word case than in the picture−word case. Yet, several studies have shown that the picture superiority effect persists on word recognition tests (Borges, Stepnowsky, & Holt, 1977; Jenkins, Neale, & Deno, 1967; Madigan, 1983; Scarborough, Gerard, & Cortese, 1979; Snodgrass & McClure, 1975). For example, Madigan (1983) presented subjects words or pictures and then tested their recognition of the same words they had studied or the word labels of the studied pictures. Despite the fact that the symbolic modality was changed between study and test for picture items, subjects still recognized items studied as pictures about 15% better than items presented as words, even though words were tested in exactly the same form as their original presentation (see the right panel of Figure 7.1).

We should also note that in free recall, as well as in word recognition, subjects must reproduce pictorial information in a form different from that in which it was studied. That is, when subjects recall a series of words and pictures, they must produce verbal responses, and in some sense, therefore, words appear to enjoy the advantage of being recalled in the same form as presented, whereas pictures should suffer from being recalled in a different form. Yet despite the potential obstacle to picture retrieval that the different study and test formats may create, recall and recognition of pictures is superior to that of words. Thus, the picture superiority effect has proved to be robust across various retrieval tasks, even though they may seem to favor word retrieval. As Madigan notes, “. . . the effects of symbolic modality (verbal versus pictorial) are large, reliable and general ones, and as such ought to influence the development of theories of cognition and memory and demand satisfactory accounting for by any such theories” (1983, p. 65). The remainder of the chapter is directed toward this end.

**Implicit Measures of Retention**

In the past 10 years psychologists investigating memory have begun to explore a new set of measures, and we are still fumbling for an appropriate categorization or taxonomy of memory in light of new data. Graf and Schacter (1985) have
provided the helpful classification of explicit and implicit measures of retention, which serves as at least a descriptive distinction. Explicit measures of retention are those in which people are asked to recollect some information; subjects know that their memories are being tested and they attempt conscious retrieval of the requested information. All the standard measures of episodic memory—free recall, cued recall, serial recall, frequency judgments, and various recognition procedures—would qualify as explicit measures of retention. Implicit measures of retention are those in which learning and retention are measured indirectly, usually through some form of repetition priming (the facilitation in processing a test stimulus because of its prior exposure, typically measured by improved accuracy or speed of responding). In these implicit tasks subjects are exposed to material (words, pictures, sentences) and then later given another task in which some form of these same items is repeated. In Table 7.1 we list the main tasks that have been used as implicit measures of memory, although the list is by no means exhaustive. Even Ebbinghaus's (1885/1964) savings method can be considered an implicit measure of retention (Roediger, 1985).

Implicit measures of retention have become of interest because they appear to reveal learning in cases where standard, explicit measures do not. For example, when patients classified as amnesic are given tests of verbal memory, their recall and recognition are notoriously poor. However, on implicit tests, such as completing fragmented words or producing the first word to come to mind to a three-letter stem (Warrington & Weiskrantz, 1970), these patients often show perfectly normal levels of priming. The implication of such results is that amnesics do not suffer difficulties in acquiring or storing information, as has long been thought; instead, the difficulty seems to be one of gaining conscious access to available information (as required on explicit tests). An ancillary implication, to state the case most forcefully, is that implicit measures of retention permit measurement of unconscious memories.

Although systematic use of such implicit measures in studying memory of normal subjects dates only to the past 10 years or less, many important discoveries have been made (see Jacoby & Witherspoon, 1982; Roediger & Blaxton, 1987, for partial reviews). Interestingly, manipulation of independent variables often

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sample references</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Savings in relearning</td>
<td>Ebbinghaus (1885/1913)</td>
</tr>
<tr>
<td>2. Reading inverted text</td>
<td>Kolers (1976)</td>
</tr>
<tr>
<td>5. Lexical decision task</td>
<td>Scarborough, Gerard, and Cortese (1979)</td>
</tr>
<tr>
<td>8. Anagram solution</td>
<td>Dominowski and Ekstrand (1967)</td>
</tr>
</tbody>
</table>
exerts different effects on implicit tasks than on explicit tasks. For example, manipulations of "elaborative coding," such as varying orienting tasks in a levels of processing experiment or forming mental images, have large effects on explicit tasks but have little or no effect on perceptual identification or word stem completion (Graf & Mandler, 1984; Jacoby & Dallas, 1981). Other variables, such as the modality and typography of studied words, which have very little effect on long-term recall and recognition, can be shown to have sizeable effects on repetition priming in implicit tasks (Graf, Shimamura, & Squire, 1985; Jacoby & Dallas, 1981; Roediger & Blaxton, 1987a). Even more impressively, some variables that exert powerful effects in one direction on recognition and recall can be shown to exert equally powerful, but opposite effects on implicit measures (Blaxton, 1985; Jacoby, 1983).

The challenge of understanding powerful dissociations among measures of retention is paramount. Various researchers have suggested that the basic underlying distinction is between tasks requiring episodic or semantic memory (Tulving, 1983), or declarative and procedural memory (Cohen & Squire, 1980), or perhaps even implicit and explicit retention (measures requiring or not requiring conscious awareness). Here we argue that the best understanding of these dissociations currently lies in the distinction made by Jacoby (1983) between conceptually-driven and data-driven tests of retention. Briefly, some tests of retention, such as free recall, provide subjects with little overt "data" to guide retrieval and so subjects must rely on conceptual processes (organization, elaboration, and the like) to provide their own cues during retrieval. On the other hand, in such tasks as perceptual identification or word fragment completion, subjects are provided some form of "data" at test, with the requirement being to produce the first appropriate response that comes to mind. In these tasks, variation in the surface features between study and test presentations produces large effects on performance; therefore, the similarities in processing the "data" are critical. The distinction between data-driven and conceptually-driven processing depends both on the similarity of the "data" provided at study and test and on task requirements. In recognition memory tasks, for example, subjects are provided with "data" in the form of copy cues, but the task requirement to decide whether or not each item is a word from the studied list causes the subject to rely mostly on conceptually driven processing (Jacoby, 1983). However, the distinct terms "data-driven" and "conceptually-driven" processing should properly be considered to represent endpoints on a continuum rather than a strict dichotomy (Roediger & Blaxton, 1987b). For example, the type of processing required in recognition can be manipulated by task demands, as shown by Johnston, Dark, and Jacoby (1985).

According to this distinction between data-driven and conceptually-driven processing, study manipulations that vary elaboration of coding (levels of processing, forming images, generating words from impoverished cues) should produce large effects on conceptually-driven tasks but little or even reverse effects on data-driven tasks. On the other hand, various methods of manipulating the physical attributes of study stimuli (changing modality, typeface, etc.) should
have large effects on data-driven tasks but little effect on conceptually-driven tasks. Roediger and Blaxton (1987b) have reviewed the body of evidence consistent with these predictions and have also elaborated the distinction between data-driven and conceptually-driven processes. Here we wish to invoke this distinction to set the stage for the current investigations. Briefly, we hypothesized that an implicit memory task that emphasized data-driven processing might reverse the usual superiority of pictures to words in tests of retention. If an exception were to be found to the usual “law” of the picture superiority effect, then perhaps it could serve as a tool to permit a new line of inquiry on picture/word effects.

Reversing the Picture Superiority Effect

The implicit, data-driven task we chose for this research was the word fragment completion task (Tulving et al., 1982). In this test subjects attempt to supply letters missing from words in order to form a complete word (e.g., _hi_b_e for “thimble”). On this test, some words have been presented during an earlier study phase and others have not. The latter set provides a baseline measure of completion to permit assessment of priming from prior study. The fragments are normed ahead of time so that each has a unique solution. Subjects taking the test in our experiments are not informed that some of the fragments are items studied earlier.

Previous research in our laboratory has shown that this test is quite sensitive to the physical format of presented information, unlike most studies of free and cued recall. For example, Roediger and Blaxton (1987b) showed greater priming of words presented visually than those presented auditorily. That is, significantly more fragments were completed when subjects had previously read the target words than if they had heard them. (Significant cross-modal priming did occur, however.) Similarly, Durgunoğlu and Roediger (1986) tested Spanish–English bilinguals on free recall and word fragment completion after they had studied a mixed list of Spanish and English words. Spanish words were remembered slightly better in free recall, probably because the subjects were Spanish-dominant bilinguals. However, on the English word fragment completion test subjects showed greater priming for words presented in English than for those studied in Spanish. In fact, no significant priming occurred in the cross-language condition. Thus, in both the cross-modal and cross-language experiments, the most priming was obtained from stimuli that most closely matched the word fragments in terms of surface features. The word fragment completion task is primarily data-driven in that it taps memory for the processing of the surface features more than the conceptual features of studied stimuli. These results led us to hypothesize that the picture superiority effect might be reversed when retention was measured by priming on the word fragment completion test, because fragment completion performance benefits most from studying physically similar stimuli, i.e., words.
Experiment 1

Our first experiment in attempting to reverse the customary picture superiority effect by using an implicit (data-driven) memory test is reported in detail elsewhere (Weldon and Roediger, 1987) and we only summarize the main points here. Subjects studied a mixed list of 42 pictures and words (half of each) presented for 5 sec per item and then, after a brief delay, received either a free-recall test or a word fragment completion test. In free-recall subjects were told to write the names of the studied pictures and words. In the fragment completion test subjects were instructed that they would see word frames and that they should complete each with a word, if possible. An example was given (\_cc\_rd\_o\_ for “accordion”). They were further told that the task was difficult and that they should continue trying throughout the 20-sec period provided for each fragment. The test was represented as a filler task to collect information for future research, with no mention made that some items had been studied previously. The fragments represented equal numbers of items studied as words and as pictures. In addition, one third of the fragments represented nonstudied items. The items were counterbalanced across the three conditions (word, picture, nonstudied) over subjects.

We expected to replicate the usual picture superiority effect for subjects who received the free recall test but hoped to reverse the effect on priming in word fragment completion. The results confirmed these expectations, as shown in Figure 7.2. On the left is free recall, showing that pictures were remembered better than words. More interestingly, depicted on the right is the amount of priming in the word fragment completion task for both pictures and words. The bars represent the amount of facilitation (priming) from studying pictures and words beyond the base rate of completing nonstudied items (.38). Study of words produced much greater priming than study of pictures, 0.26–0.07. Thus, on the implicit memory test of word fragment completion, we reversed the usual picture superiority effect.

A critic might complain that our reversal of the picture superiority effect under these conditions was uninteresting because we merely contrived conditions that artificially favored word retrieval. That is, the word fragments provided at test obviously matched the stimuli studied as words better than those studied as pictures. It is therefore no surprise that word fragment completion reveals greater priming from words than pictures. However, this argument from our hypothetical critic evaporates when one considers the recognition memory results shown above on the right in Figure 7.1 (e.g., Madigan, 1983), where presentation of pictures produced superior recognition to presentation of words even when the recognition test was wholly composed of words. In such a recognition test the overlap between the surface features of study and test items is even

*All results described in the text were statistically significant at or beyond the .05 level of confidence by conventional tests, unless otherwise noted.
Figure 7.2. The picture superiority effect in free recall (left panel) was reversed when retention was measured by priming on the word fragment completion task (right panel). Priming is computed by subtracting the baseline completion rate for nonstudied items from the completion rate for studied items for each subject. Data from Weldon and Roediger (1987, Experiment 1).

greater than in the word fragment completion test, and yet the picture superiority effect still obtains. Although recognition may have what we are calling a data-driven component, under “standard” conditions in which items in recognition tests are paced slowly, recognition behaves largely as a conceptually-driven test (Jacoby, 1983; Johnston et al., 1985).

The reversal of the picture superiority effect we found in word fragment completion priming is therefore genuine and also helps point up a critical feature of the distinction between data-driven and conceptually-driven tasks: these tasks do not necessarily differ in the amount and kind of data presented at test but differ instead in terms of the processing requirements of the test. Conceptually-driven tasks require subjects to reflect on the meaning (or associations or images or elaborations) of a concept, whereas data-driven processing usually requires a response to the presented data without the necessity for higher level reflection.

Experiment 2

A potential problem with the interpretation of our results occurred to us. We obtained the picture superiority effect in free recall, whereas our reversal was obtained in the word fragment completion task. The difference we obtained between free recall and word fragment completion may have been resulted simply from different cueing properties in the two situations, rather than to differences in types of processing. That is, words were retained better than pictures in the word fragment completion task (a cued task), but pictures were retained better than words in free recall. Perhaps it is the case that words are simply remembered better than pictures on tests that involve cued recall. Of course, the picture
superiority effect in word recognition (Madigan, 1983) would seem to argue against this possibility. However, we decided to test it directly by seeing if we could obtain both picture superiority and word superiority under identical cuing conditions by varying only the instructions telling subjects how to use the cues. We hoped to find that subjects given Implicit instructions would reveal greater priming for words than pictures, whereas those given Explicit instructions would remember pictures better than words.

In this experiment all subjects studied mixed lists of 26 pictures and 26 words for 5 sec each then received word fragments at test. Thirteen test fragments were nonstudied items, and all items were rotated so that they served as pictures, words, or nonstudied items an equal number of times across subjects in each test condition. The critical manipulation was the instructions given at test. Twenty-five subjects received Implicit retention instructions; i.e., they were told to try to solve each fragment with the first word that came to mind. They were told that they were helping prepare materials for future research and were not told that some of the fragments were items from the study list. This condition replicated the word fragment completion test condition in Experiment 1. After they finished the fragment completions, these subjects were given 7 min to free recall the pictures and words. The other half of the subjects received the same word fragments in the same order at test but were given Explicit retrieval instructions; i.e., they were told to use the fragments as retrieval cues to help them remember the pictures and words they had studied. All subjects were given 15 sec to complete each fragment. The manipulation of instructions permitted us to equate the cues in the implicit and explicit retrieval conditions to determine whether the presentation of such cues always leads to superior retention of words relative to pictures.

The fragment completion results are displayed in the left panel of Figure 7.3. First, notice that the reversal of the picture superiority effect reported in the first

![Diagram](image-url)

**Figure 7.3.** The picture superiority effect was reversed on the implicit word fragment completion test (left panel), but pictures and words were retrieved equally well when the same cues were used on an explicit task (right panel).
experiment was replicated in the group that received Implicit fragment completion instructions: words produced significantly more priming than pictures. The baseline rate of completion was 0.32, and the total completion rates for words and pictures were 0.54 and 0.42, respectively. Interestingly, in this group the usual picture superiority effect was obtained on the free-recall test given after fragment completion. Proportions recalled were 0.37 for pictures and 0.26 for words. This picture superiority effect in free recall occurred despite the fact that subjects had received a second study opportunity during the fragment completion test for more of the original word items than picture items.

In the group that received the same fragment cues with the Explicit cued-recall instructions, the better performance obtained from studying words than pictures under Implicit instructions was eliminated. As shown in Figure 7.3, there was no significant difference between the number of pictures and words recalled. Although Explicit instructions altered the pattern of results, we had expected cued recall to produce a picture superiority effect. One possible—although speculative—interpretation of our failure to find the picture superiority effect in explicit cued recall with word fragment cues is that there is simply a strong and unavoidable data-driven component to such cues. That is, when impoverished data are provided to subjects, it may be difficult or impossible to override data-driven processing completely by explicit memory instructions. In fact, Blaxton (1985) developed a graphemic cued-recall test in which subjects were given words that looked and sounded like target words (for example, the word “chopper” as a cue for the target word “copper”). She showed that under explicit memory instructions, her graphemic cued recall task behaved like a data-driven task, because it was sensitive to the physical format of studied information but relatively unaffected by conceptual manipulations at study. Therefore, our word fragment cued-recall task, like Blaxton’s graphemic cued-recall test, might have an inherently data-driven aspect.

Experiment 3

If the preceding reasoning were correct, then we failed to find the picture superiority effect in cued-recall in Experiment 2 because the cues we used (word fragments) had an inherent data-driven component that favored word recall. One way to test this idea was to develop another cued-recall task in which the cues would provide only a conceptual relation to the target words. Under these conditions, the picture superiority effect should emerge. Surprisingly, we could find no experiment appropriate to our needs in the literature, so we conducted a third experiment in which we provided extralist cues that shared no physical “data” with the targets. Instead, the cues were strong associates of the targets (e.g., “emergency” as a cue for “ambulance”) with the latter concept presented either as a picture or a word in the study list. Thus, the cued recall task was conceptually-driven, because subjects were required to rely on associative information in order to use the cues to recall the targets. If our hypothesis is correct, i.e., that the word fragment completion task is too data-driven to permit a picture
superiority effect under explicit cued-recall instructions, then removing the data-driven component of the cues should restore the picture superiority effect.

Forty subjects studied mixed lists of 30 pictures and 30 words, with study format counterbalanced across subjects. They then received the extralist cued-recall test, which contained 60 cue words associatively related to the studied items. Subjects were told to use these cues to help them remember the pictures and words, and to write each recalled item next to the cue. They were given 15 sec to attempt to recall an item for each cue.

Cued recall was significantly higher for pictures than words, with recall proportions of 0.72 for pictures and 0.56 for words. Thus, the picture superiority effect was obtained in cued recall with extralist associates as cues. This result suggests that the failure to find a picture superiority effect using word fragments as cues in the previous experiment probably owes to the inherently data-driven nature of the retrieval task, and not to any general absence of picture superiority effects in cued recall. Cued recall with associatively related items is a conceptually-driven task because it requires the use of information about an item's meaning and semantic relationships with other items. Therefore, because pictures induce more conceptual processing, pictures were remembered better than words on this cued recall test.

We emphasize that our hypothesis—that cued recall with word fragment cues is inherently data driven—is speculative and admittedly ad hoc at this point. A reasonable objection to the suggestion that explicit cued recall with word fragments is data-driven is that we also maintain that explicit word recognition is conceptually driven, and yet the word recognition test provides more data than a fragment completion test. One might argue that the recognition test should therefore be more data-driven than the cued fragment completion test with explicit memory instructions. However, the critical factor determining the degree to which a task is data-driven may be the degree to which it is data limited. Performance on data-limited tasks such as perceptual identification and word fragment completion may depend more heavily on recapitulation of perceptual rather than conceptual processes. These ideas require further research but open exciting possibilities for understanding the mechanisms underlying the distinction between conceptually and data-driven tasks.

Experiment 4

In the final part of this section on our own research, we raise one last question concerning our demonstration of a reversal of the usual picture superiority effect on priming in word fragment completion. One possible description of most of our results so far is that the picture superiority effect is obtained on explicit memory tests (free recall and extralist cued recall), but not on implicit memory tests (word fragment completion). Perhaps the explicit/implicit distinction provides at least a description of when the picture superiority effect will be found or reversed. On the other hand, we have argued that the proper distinction underlying our results and those of others is that between data-driven and conceptually-
driven processing. A natural confounding has been built into most research, such that almost all explicit memory tests are what we would call conceptually-driven and almost all implicit memory tests (see Table 7.1) are data-driven. Blaxton (1985) has performed experiments in which she has developed variants on the standard tasks (that is, she developed a data-driven explicit task and a conceptually-driven implicit task); she has shown that the requirements of the task in terms of its data-driven or conceptually-driven component was the critically important feature, and not whether the task was implicit or explicit. We used the same logic in an experiment in which we sought to show that the picture superiority effect could be obtained or reversed on an implicit memory task depending on the type of data presented at test.

Subjects in this experiment studied a mixed list of words and pictures and then received one of two types of implicit memory tests (Weldon & Roediger, 1987, Exp. 4). One test was the word fragment completion task in which fragments representing studied items (pictures and words) were intermixed with fragments of nonstudied items. Other subjects took a newly developed picture fragment identification test in which their task was to name severely degraded pictures created by eliminating critical features. (These picture fragments were normed; several examples with their verbal labels appear in Figure 7.4.) As in word fragment completion, subjects taking the picture identification test were asked to identify fragments of items previously presented as pictures or words, as well as nonstudied items. As usual, subjects taking both types of test were given instructions to complete the word fragment or identify the picture fragment as quickly as possible, with no mention of the fact that some items had been studied previously.

Both tasks used in this experiment are implicit memory tasks, but the type of data provided should match features of the study episodes in different ways. The “data” in the word fragments match those in studied words much better than
Figure 7.5. A picture superiority effect was found in the picture identification task, but greater priming (studied vs. nonstudied completion rates) occurred for words than pictures on the word fragment completion test.

... those in pictures, but the converse was true of the picture fragment identification task. Here, match of the test stimuli was better with studied pictures than with studied words. Thus, despite the fact that both are implicit memory tasks, we predicted a picture superiority effect for the picture fragment identification task, but greater priming from words on the word fragment completion test (as in Figure 7.2).

The results followed our expectations quite well, as can be seen in Figure 7.5. The picture identification test did lead to greater priming from prior study of pictures than words, whereas words produced greater priming than pictures on the word fragment completion test. (Note that small cross-form priming was obtained on both tests, although it was reliable only on the word fragment completion test.) Because the picture superiority effect can be obtained or reversed on implicit memory tests, depending on the type of data provided at test, we conclude that a possible generalization from our earlier work—i.e., that picture superiority occurs only on explicit tests but is reversed on implicit tests—is inaccurate. (See Warren and Morton, 1982, for a similar demonstration with perceptual identification.) Instead, the critical determinant is the degree to which the retrieval test accesses the record of the type of processing performed during the study episode.

To summarize the main results and conclusions emerging from our experiments:

1. The picture superiority effect can be reversed by changing retrieval demands alone.
2. Fragment completion tests such as those used here are primarily data driven: that is, they tap memory for the processing of stimulus surface features.
3. Whereas conceptually-driven tests typically produce better retention for pictures than words, data-driven tests produce superior retention for those stimuli whose physical attributes most closely match those of the test stimuli.
Related Research

Our results above showed that words are retained better than pictures when memory is measured by an implicit, data-driven memory task of completing word fragments. Of course, our demonstration is limited to just this one measure—priming in word fragment completion. It would be comforting to know that our claims generalize to other tasks. Fortunately, some relevant evidence that converges with our conclusion does exist. However, the researchers who collected it were mostly concerned with other issues and did not consider their findings to show a reversal of the picture superiority effect, probably because the tasks used were considered to measure perception rather than memory. However, the work of Jacoby, Kolers, and others has convinced most researchers that priming in “perceptual” tasks can provide important information about memory (see Kolers & Roediger, 1984), so this evidence is relevant in the present context.

Wilma Winnick and Stephen Daniel (1970) reported what should properly be considered a landmark study. However, because of the language in which they cast their results, and perhaps because of the temper of the times, their report went largely unnoticed for a decade or more. Their Experiment 2 was remarkable because it clearly showed (a) the generation effect in free recall (Slamecka & Graf, 1978), (b) the reversal of the generation effect on a perceptual fluency test of implicit memory (Jacoby, 1983), (c) the picture superiority effect in free recall, and (d) reversal of the picture superiority effect on an implicit test. The latter two points are most relevant to present concerns, but we shall describe the experiment in its entirety to help resurrect it.

Winnick and Daniel (1970, Exp. 2) had subjects read four words aloud from each of three types of stimulus display, which instantiated one independent variable in their experiment. Subjects said the word when presented with (a) the word itself (e.g., “airplane”), (b) a picture of the named object, or (c) a definition of the object. Four other words from the same set were not presented but were used as a baseline for one of the tests. Items were counterbalanced across conditions over subjects and the three conditions of word naming were realized within subjects.

The other primary factor was the type of test given. Half the subjects received a free-recall test for the stimuli, whereas the other half were given a tachistoscopic word identification test. In the latter, the threshold for naming each word was measured by the ascending method of limits beginning with an exposure duration of 10 msec and increasing by 5 msec increments on each trial until subjects could identify the word. The measure of interest was the visual duration threshold, or the average amount of time subjects needed to identify words in the different conditions. The nonstudied control words were used as a baseline against which to measure priming in the various study conditions. Subjects were not informed that the studied words were to be presented during the threshold test, so this measure qualifies as an implicit measure of retention.

The results of Winnick and Daniel’s experiment are shown in Figure 7.6. Consider first the free-recall results in the left panel. Labels that subjects generated
in response to a picture or a definition were better retained than studied words, revealing the picture superiority effect and the generation effect. The comparison of recall of words generated in response to pictures and to definitions raises an interesting thought concerning the picture superiority effect: perhaps the advantage in recall of pictures to words is not caused by some form of imaginal coding, which is the usual account, but should be considered instead a species of the generation effect. That is, good recall or recognition of words generated from pictures might be achieved by the same mechanisms producing the generation effect and not by dual coding in imaginal and verbal systems (Paivio, 1971). Words generated in response to definitions can be conceived as a relevant comparison condition for the usual picture presentation condition, because we may assume that processing occurs wholly within the verbal system. In this control case, recall of the generated responses is actually slightly superior to recall of words generated to pictures. The possibility that the picture superiority effect in recall and recognition is really a result of generation (rather than imaginal coding) is further discussed by Weldon and Roediger (1987) in light of new data on the issue. The main point to be taken from the left panel of Figure 7.6 for current purposes is that Winnick and Daniel did find superior free recall of pictures relative to words.

The right panel of Figure 7.6 shows the amount of priming on the tachistoscopic threshold measure. Priming here indexes the reduction in word recognition thresholds for the two study conditions relative to nonstudied words. On this measure, studied words showed considerable priming, but items gener-
ated from pictures or definitions did not. (Although the amount of priming for words may seem small—a 13-msec advantage over control words—this actually reflects considerable priming in the threshold task. Thresholds averaged about 62 msec in the control condition, so the priming from words represents about a 20% improvement.)

The overall pattern of results thus reflects a strong dissociation: the effect of the various study conditions on free recall was exactly reversed on the implicit threshold identification task (cf. Jacoby, 1983; also Blaxton, 1985). Pictures were remembered better than words on free recall, but words were retained better than pictures as measured by priming on the identification task. Thus the results anticipate our own shown in Figure 7.2 and we would argue that they support our reasoning: free recall is a conceptually-driven test, whereas tachistoscopic identification is data-driven. Of course, Winnick and Daniel (1970) did not discuss their results in any of the terms used here—most of which were not in use in 1970—but instead discussed them in terms of the response availability and perceptual sensitization created by various conditions.

Another important aspect of the Winnick and Daniel results—noted rather belatedly by Morton (1979)—is that they call into question his logogen model. According to the original version of the model, logogens are abstract representations of words or concepts activated by any relevant source. The firing of a logogen leaves residual activation that then produces priming if the same stimulus is repeated after a short delay. Because the logogen is abstract, activation and priming should occur whenever the logogen is activated and by whatever means. However, the Winnick and Daniel (1970) results showed that priming occurred when subjects had previously seen words, but not when the equivalent concepts were generated from pictures or definitions. Morton (1979; Clarke & Morton, 1983) replicated Winnick's and Daniel's (1970) findings and extended them in important ways, causing changes in the logogen model, too (see also Jackson & Morton, 1984; Warren & Morton, 1982). Although providing details of this work is beyond the scope of this chapter, in general the results are consistent with the notion that priming in either visual or auditory identification or threshold tasks behaves in a “data-driven” fashion.

Another relevant set of experiments was reported by Scarborough et al. (1979). In their experiments subjects generated words in response either to the name of the word or to a picture representing the concept. A third set of control words was not presented. Later subjects saw the words generated at study presented in a lexical decision task (Experiments 1 and 2) or in a standard recognition memory procedure (Experiment 3). In the lexical decision task, words were mixed with nonwords, and subjects made word/nonword judgments about each item; in the recognition experiment words were mixed with distractors and subjects made old/new decisions about each item. The measure of interest in the lexical decision task was repetition priming, measured as the increased speed with which previously generated items were identified as words, relative to the nonstudied items. The measure of interest in the recognition memory task was, of course, the ability to discriminate old from new items as represented by $d'$. 
7. Reversing the Picture Superiority Effect

![Graphs showing recognition and repetition priming](image)

**Figure 7.7.** Results from Scarborough, Gerard, & Cortese (1979). Pictures were better recognized than words on a word recognition test (left panel; data from Experiment 3), but repetition priming (facilitation in decision times) in the lexical decision task was greater for items previously studied as words rather than pictures (right panel; data from Experiment 2). Only results from low-frequency items are given.

The results are presented in Figure 7.7 for low-frequency words. In the left panel are $d'$ values and the usual finding was obtained: words generated in response to pictures were better recognized than those actually seen as words, even on a word recognition test. However, considering repetition priming in the lexical decision task as an implicit measure of retention, the picture superiority effect was reversed (see the right panel). Only a nonsignificant 5-msec priming effect was found for words generated in response to pictures, but prior naming of the word itself produced a 50-msec priming effect. Although Scarborough et al. (1979) did not explicitly note that they had reversed the picture superiority effect on some form of memory test, their interpretation of these results is generally similar to the one offered here. Somewhat curiously, the effects in the lexical decision task were not well replicated in one experiment with medium- and high-frequency words. However, Durso and Johnson (1979) also found greater repetition priming from words than from pictures on tasks requiring either naming or categorization of words. Therefore, the effect seems genuine.

Finally, we should note one other relevant experiment, albeit briefly. Kroll and Potter (1984, Exp. 5) employed a decision task in which subjects had to categorize pictures and words as representing “real objects” or not when pictures and names of objects were mixed with nonwords and nonsense “objects.” The concepts representing pictures and words were sometimes repeated, and then either in the same form or a different form. In general, Kroll and Potter found strong repetition effects. When the repetition occurred in the same surface form (picture–picture or word–word), but either no (word to picture) or slight (picture to word) priming was obtained with cross-form repetitions. In general, the results agree that these “priming” tasks are largely data-driven (enhanced by matching
surface features) and little affected by conceptual processing (a correspondence in the underlying concepts).

Synopsis and Extensions

The superiority of pictures to words in retention has been considered as a general, established fact, as indeed it is in most episodic memory tests. However, the results from our experiments (Weldon & Roediger, 1987, and from others reviewed in the preceding section, show that on several implicit measures of retention, words produce superior performance (priming) relative to pictures. We suggested that the reason for the picture superiority effect on most explicit measures of retention is that these depend on conceptual information (elaborations, associations, images and the like) for good performance; in short, they are conceptually-driven. On the other hand, most implicit measures of retention require subjects to produce the first accessible response given some data-limited presentation. In these cases the match between surface features of the test items and the items studied earlier determines how good performance will be, with conceptual factors (elaboration, etc.) playing little or no role. Thus, when test items consist of words on these data-driven implicit tests, priming from prior study of words exceeds priming from prior study of pictures, thus reversing the usual advantage of pictures to words in tests of retention. Of course, if the “data” in the test are pictorial, then the picture superiority effect will still be found on data-driven tests (see Figure 7.5).

What implications do our results with data-driven tests, showing reversals of the usual picture superiority effect, have for theories of imagery in remembering? First, we believe that our results do point up a critical oversight in imagery research (and most other memory research) by neglecting task differences in remembering (Jenkins, 1979). Our results show that, on appropriate tests, words are retained better than are pictures and so there is nothing inherently more memorable—as a main effect across every retention test—in pictures than words. However, the implication of our results for specific theories of imagery in remembering may be less profound. In Paivio’s (1971) dual code theory, our results could be interpreted as showing test conditions in which verbal coding can be more powerful than imaginal encoding. However, two codes would still be postulated. In terms of Nelson’s (1979) semantic sensory model, implicit memory tests may tap the record of sensory features, whereas conceptually-driven tests would tap the semantic component. At this general level, our results would not be inconsistent with the main postulates of either theory.

The interpretation of our data offered above can be put into the more general terms of transfer-appropriate processing, in the various forms in which that idea has been raised (Morris, Bransford, & Franks, 1977; see also Kolers & Roediger, 1984; McDaniel, Friedman, & Bourne, 1978; Stein, 1978). The basic notion is that learning activities may be of many kinds (offering many possibilities for encoding material) and that performance on some test of retention or transfer
will benefit to the extent that the kind of processing or the procedures required by the test recapitulate those of the original learning experience. (Many similar ideas exist in the psychology of learning and memory—stimulus generalization, the encoding specificity hypothesis—but we prefer the variants mentioned above because the emphasis is on a matching of processes or procedures rather than a matching of mental elements, structures, or contents.) Put in these broad terms, our argument is that most explicit, conceptually-driven retention tests favor processing of pictures because the procedures in picture encoding are conceptually richer. Similarly, for implicit tests that are data-driven and involve rapid processing of words, prior study of words will produce more priming than prior study of pictures.

The concept of transfer-appropriate processing is, we believe, fundamental to a proper understanding of learning and memory. Roediger and Blaxton (1987b) discuss these issues in more detail. We shall illustrate the applicability in the present context by reference to other work on imagery and mnemonics. Mnemonic devices have long been known to produce very high levels of recall on certain types of memory test (e.g., Bellezza, 1981; Bower, 1970; Yates, 1966). Critics of mnemonics have argued that their use is really quite limited, involving situations where people remember lists or series of elements. Another way of stating this criticism is that traditional mnemonic devices may provide appropriate transfer for only a few measures of retention, and for limited materials. That is, the encoding activities encouraged by mnemonics may require (even demand) appropriate forms of test to reveal their benefit and be useless for most practical purposes. However, we believe that suitable mnemonics can improve performance in many relevant situations, especially when people are required to recall information with little aid from external cues. Most common mnemonics provide the means of supplying a good set of self-generated cues.

An example of transfer-appropriate processing induced by mnemonics occurs in a comparison of the effectiveness of mnemonics in which different measures of recall were obtained. Roediger (1980) instructed subjects to use one of five mnemonic techniques to study and recall lists of 20 words. The techniques were (a) imagery—subjects were to imagine the referent of each word; (b) the link method—subjects were to imagine the object represented by each word and link it to the next; (c) the familiar method of loci; (d) a peg method involving a number–word rhyme scheme ("one is a gun," etc.); (e) a rehearsal control group who repeated each word until the next word was presented. After a small amount of training, subjects learned and recalled successively three lists of 20 words. On their test sheets they were given the numbers 1 to 20 in a column and asked to write down the appropriate word beside each number. However, they were further told that they should be sure to write down all words they could remember, either guessing at the proper locations when unsure or writing them at the bottom of the page.

The results were analyzed in terms of two measures. First, the total number of items recalled in any position was measured and is presented in the solid columns of Figure 7.8, with the bars representing averages across three lists. All four
groups that used imagery recalled significantly more items than did the Rehearsal control group. Differences among the four groups using imagery were small, although the Link and Loci groups recalled significantly more words than the Imagery group (see Roediger, 1980 for further details). The situation is rather different when subjects’ data were rescored by a strict positional criterion. By this measure, subjects were given credit for correctly recalling a word only when it appeared on their recall sheets next to the correct number. The results are shown in the open bars of Figure 7.8, where differences among conditions are much greater. Recall was low and did not differ between the Rehearsal and the Imagery groups, was better in the Link group, and was best in the Loci and Peg groups. Of course, in some sense these results are not surprising, because the Peg and Loci conditions encourage encoding of order information, whereas the simple Imagery and Rehearsal instructions do not. The link mnemonic encourages some attention to a fixed order, but not as much as the loci and peg systems.

The important point is that to see marked differences in effectiveness of the various mnemonics, one must test knowledge in an appropriate way. Simply by looking at the total number of items recalled, one might conclude that the mnemonics had little advantage over a simple instruction to form separate images (or to rehearse, for that matter). However, when order is taken into account, the peg and loci mnemonic techniques showed over a 100% advantage relative to the Imagery or Rehearsal control groups. One could imagine a further test that would show a marked advantage for only one of the five groups, although we know of no study documenting this point. Suppose subjects had been asked to produce the words associated with the various numbers when the numbers were given in a random order (what was the fourteenth word? the eighth? etc.)? Certainly one would predict the Peg group to perform much better than the other groups in
terms of their speed and accuracy of response, because their encoding strategy uniquely promoted retention of this sort of information. Once again, appropriate retention tests may reveal transfer of the specific information learned, where inappropriate tests do not.

From the consideration of transfer-appropriate processing, successful mnemonics can probably be constructed for any retention test. One needs to ask, what are the critical requirements of the test or evaluation procedure, and what kind of learning activities would promote (transfer to) the particular test? An excellent example of this logic is given by Rohwer and Thomas (Chapter 20, this volume) in which they discuss the role of mnemonic strategies in promoting effective study in classroom settings. Traditional mnemonics such as the peg method, may be of little use in most educational applications. However, by considering the various task demands, learning materials, subject characteristics, and learning strategies (Jenkins, 1979), one can develop efficient learning strategies with important practical applications (see Figure 20.1 in Rohwer and Thomas, Chapter 20, this volume). McDaniel and Pressley (1984) showed how the keyword method could be used to important benefit in an educational situation when put in an appropriate context. The concept of transfer-appropriate processing thus serves as a useful theoretical context for understanding many phenomena of learning and memory, and also serves the practical function of potentially guiding discovery of more efficient methods of instruction in educational and other applied situations.

Acknowledgments. This research was supported by NIH Grant RO1 HD15054 awarded by the National Institute for Child Health and Human Development.

References


