MEMORY AND LEARNING
The Ebbinghaus Centennial Conference

edited by
DAVID S. GORFEIN
ROBERT R. HOFFMAN
Adelphi University

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1987 Hillsdale, New Jersey Hove and London
Rereading Über das Gedächtnis reminds one of Ebbinghaus’ many contributions: his clever methods, his careful (even compulsive) experimentation, his remarkably systematic findings, and his crisp, economical writing style. However, a contemporary reading leads to the more surprising insight that not only did Ebbinghaus make the many discoveries for which he is justly celebrated, but that other, overlooked, passages also seem important today (Hoffman et al., this volume; Roediger, 1985; Slamecka, 1985).

One such surprise is that Ebbinghaus clearly distinguished among different forms of remembering. The idea that there are various modes of expressing retention has received little attention by experimental psychologists in the 100 years since publication of his book, except for the relation between recall and recognition. However, this issue has become a prime concern for some contemporary workers and is the focus of the research to be reported in this chapter.

In the first two pages of his 1885 monograph, Ebbinghaus distinguished between voluntary and involuntary remembering. In the former, one must make an effort at recollecting the events in question. In the latter, the memories of the events come to the rememberer unbidden, or “automatically.” Furthermore, and importantly for present concerns, Ebbinghaus noted that involuntarily produced memories may or may not carry with them any direct feeling of familiarity. Sometimes we are aware that these involuntarily produced states of mind refer to memories of past experience through such feelings of familiarity, but on other occasions we can only infer this indirectly. Finally, Ebbinghaus pointed up a third possibility and he reckoned that it represented a “large group” of memories. These are memories that “remain concealed from consciousness and yet
produce an effect which is significant and which authenticates their previous experience” (1885/1964, p. 2).

The use of what Ebbinghaus (p.8) called introspection—or measures of recall or recognition, as we would say today—fails to capture these distinctions. Almost all such memories produced by the rememberer are of the voluntary kind. Perhaps a few are involuntary; sometimes memories seem to occur spontaneously during lengthy repeated tests for reasons that subjects cannot articulate (e.g., Erdelyi & Becker, 1974; Roediger & Payne, 1982), and occasionally an event seems familiar but no specific experience can be identified as giving rise to this feeling. However, Ebbinghaus’ third class of “unconscious retention” must, by definition, escape notice by methods of recall, recognition, or similarly derived measures that depend on conscious recollection (e.g., feelings of knowing, frequency judgments).

Ebbinghaus’ relearning and savings methods overcame some of the drawbacks of introspection because savings for material could be shown even when it could not be voluntarily reproduced. As Hilgard (1964) remarked in his introduction to the Dover reprint of On memory:

Ebbinghaus took the fact of ease of relearning something once known—a fact so plausible that it must have been often observed—and made it part of science by developing the quantitative saving score, in which the saving in relearning is scored as a per cent of the time (or trials) required in original learning. This is a genuine scientific advance, although once accomplished it is so easily comprehended that one wonders why nobody else thought of it. (p. viii)

Today we might wonder why so few researchers use it.

In analysis of one experiment (Chapter 5), Ebbinghaus asked whether the ability to recollect a series played a role in relearning. He relearned some lists that had previously been learned very well and that he had recognized during relearning, as well as other lists that had not been so practiced and which he did not recognize during relearning. Although the overlearned series were more easily relearned, he found no evidence in the savings scores to indicate that conscious recollection affected relearning in any way. That is, no discontinuity occurred in the savings function for the overlearned lists that were recognized during relearning. Although the issue was not explored systematically, this finding might be said to be the first experimental demonstration of dissociation between states of awareness and performance (Tulving, 1985a), the topic of the present chapter.

Savings in relearning is a useful measure, for it can potentially show retention even when the subject lacks conscious knowledge of the material (Kolers, 1976; Nelson, 1978). However, without additional measures the savings score cannot differentiate among the various forms of retention that Ebbinghaus described. Most modern researchers approach the issue by examining performance across
different tasks, only some of which require conscious awareness of the learning episode for successful performance (e.g., Jacoby & Dallas, 1981).

The aim of the present chapter is to discuss dissociations between measures of conscious recollection (recall and recognition) and other measures of transfer or priming in which conscious recollection of studied events is not necessary for successful performance. The purposes of the present chapter are to (a) add some information to the empirical data base showing these dissociations, and (b) provide observations that, in our opinion, bear strongly on their proper explanation. In the next section, we review selected studies that are critical to our own experimental research. To anticipate, we examine the effects of presenting information in various surface forms (such as visual or auditory, or drawing or word) on retention as measured in various ways. In the third major section, we discuss theoretical accounts of the empirical findings and suggest that one approach is currently more promising than its competitors as an explanation of dissociations among measures of retention. We conclude with a few remarks about this approach and suggest some directions for future research.

MODES OF KNOWING

The concern with forms of knowing and performance has been raised in many papers over the past few years, and only a few studies will be described here to set the stage for the present research (see Jacoby & Witherspoon, 1982; Kolers & Roediger, 1984, for more extensive reviews). Of primary interest, as with some parts of Ebbinghaus’ work, is the relation between consciousness and performance. Modern work on this problem springs from at least two related lines of work. First, studies of densely amnesic patients have shown that although they perform dismally in recalling or recognizing recently presented information, their performance on more subtle transfer or priming measures indicates that they have processed and stored the information. For example, Warrington and Weiskrantz (1970) compared performance of amnesic patients with normal patients at the retention of word lists. The amnesics performed much worse than normals when tested for recall or recognition of recently presented information. However, they also tested patients with another task in which they had to name severely degraded words that could not be identified unless recently presented. The amnesic subjects exhibited normal amounts of priming when they were given fragmented words and asked to name them.¹

¹Throughout this chapter priming refers to the benefit on a task accruing from presentation of an item. For example, in a task of completing fragmented words, if subjects correctly identify 50% of the items after studying them and the completion rate for the nonstudied words is 30%, then the priming effect is 20%. Priming refers to the difference between studied items and nonstudied (unprimed) items on a task.
Many other studies have confirmed the conclusion that amnesics show performance just as good as that of normals, even when tested on verbal materials, as long as the test does not require conscious recollection (e.g., Graf, Squire, & Mandler, 1984; Jacoby & Witherspoon, 1982; Shimamura & Squire, 1984). These patients then obviously have no difficulty encoding or storing some type of representation of the material; the difficulty is in gaining conscious access to it. Put another way, they are “amnesic” on only some tests of retention, apparently those involving conscious recollection. These studies show that such patients’ knowledge of what they know can be dissociated from their actual performance.

These phenomena might only be regarded as curiosities to researchers interested in normal memory functioning if it were not for a related line of research that is also relevant to present concerns. This second line of research demonstrates in normal adults the same phenomena found in amnesics. That is, normal subjects also show dissociations between conscious recollection (as assessed by recall or recognition) and retention as measured in ways that do not require conscious awareness of the prior learning experience. For example, Kolers (1976) required subjects to read passages presented in an unusual inverted typography until they became proficient at it. A year later the subjects were unexpectedly retested and asked to read the same inverted passages intermixed with new ones from the same source. Subjects showed savings in reading speeds for the previously read passages relative to the new ones and, further, these benefits were uncorrelated with recognition judgments as to whether or not the sentences were ones they had previously read.

Many similar dissociations have been reported by Larry Jacoby and his colleagues (e.g., Jacoby, 1983; Jacoby & Dallas, 1981; Jacoby & Witherspoon, 1982) in memory for presentations of individual words. Typically, subjects in the experiments were presented with words under various study conditions and then tested on one of two tests. Some subjects were given a standard yes/no recognition test in which old and new words were mixed together and the task was to pick out the previously studied words. The other test was one of perceptual identification in which words were tachistoscopically presented for about 30 milliseconds and the subjects' task was simply to read them aloud. In the latter task, the dependent measure of interest is priming, or how much benefit having recently studied the words produces in naming them, relative to the case when they were not previously studied. Jacoby and Dallas (1981) showed in their Experiment 1 that when words were presented with questions entailing attention to their graphemic, phonemic, or semantic features, a standard “levels of processing effect” was found in recognition. Semantic encoding produced the best

\footnote{Jacoby and Dallas (1981) refer to this task as perceptual recognition, but here we prefer to call it perceptual identification so as to better distinguish it from recognition memory in describing the experiments.}
recognition, followed by phonemic and then graphemic encoding. However, on
the perceptual identification test these conditions produced equal amounts of
facilitation relative to nonpresented control words. This pattern, like that of
Kolers (1976), represents a dissociation between performance on a test of con-
scious recollection (recognition) and performance on a second test in which
conscious recollection is not required, in this case perceptual identification.

An even more dramatic dissociation is found in Jacoby’s (1983) experiments
in which subjects studied antonyms, such as cold, as target items in one of three
contexts. The context in which the antonyms were placed caused them to be
processed in three different ways across conditions. In the No Context Condition
three Xs appeared to warn subjects that an item was about to be presented; in the
Context condition the antonym (hot) preceded the target word; and in the Gen-
erate condition, subjects saw hot followed by three question marks and had to
generate cold. In all three study conditions, subjects read the context word (if
any) silently and spoke the target word aloud.

Two different forms of test were given, which produced opposite patterns of
results, as shown in Table 24.1. In a recognition test, subjects performed best in
the Generate condition, next best in the Context condition and worst in the No
Context condition. This pattern replicates the “generation effect” (Slieman &
Graf, 1978) in that generated items are remembered better than those read in the
No Context condition. However, on the perceptual identification test, in which
words were flashed and the subjects’ task was simply to name them, the pattern
of results was exactly opposite that of recognition. Note that No Context items
were best identified and Generate items identified least often. In fact, across
Jacoby’s (1983) experiments the Generate items showed little (Experiments 2

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Yes/No Recognition</th>
<th>Perceptual Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Context (XXX–COLD)</td>
<td>0.56</td>
<td>0.82</td>
</tr>
<tr>
<td>Context (HOT–COLD)</td>
<td>0.72</td>
<td>0.75</td>
</tr>
<tr>
<td>Generate (HOT–???)</td>
<td>0.78</td>
<td>0.67</td>
</tr>
<tr>
<td>Nonstudied</td>
<td>—</td>
<td>0.60</td>
</tr>
</tbody>
</table>

(Reproduced with permission.)
and 3) or no priming (Experiment 1) relative to nonstudied items. The generation effect is thus test-dependent.

We should briefly summarize our interpretation of Jacoby’s (1983) results in order to introduce concepts that we use to describe our own results. Jacoby’s findings can be interpreted as illustrating the principle of “transfer appropriate processing”: Different study conditions and types of test require various kinds of processing, and performance on a test will depend on the overlap between the type of processing engendered by the encoding condition and that required by the test (Morris, Bransford, & Franks, 1977). In Jacoby’s (1983) terms, reading the word without context (xxx-COLD) involves data-driven (bottom-up) processing, whereas generating a word when given its antonym (hot—???) requires conceptually driven (top-down) processing. That is, in the No Context condition there is no other means for the person to produce cold than for the data (the letters forming the word) to be “driven through” the cognitive system, bottom-up. In the Generate condition, on the other hand, the visual features (the letters) specifying the response cold are absent, and it must be produced by inference from the related concept hot and the rule to produce opposites. Presumably, subjects in the Context condition (hot—COLD) used a mixture of these two forms of processing. With regard to the tests, recognition memory is assumed to depend heavily on conceptually driven processing, whereas perceptual identification is assumed to require data-driven processing (Jacoby, 1983).

Given these assumptions, the pattern of results in Table 24.1 can be explained. Items produced during study in the Generate condition required the greatest amounts of conceptually driven processing, so they should be better recognized than items in the other two conditions, as was the case. On the other hand, items produced in the No Context condition required data-driven processing and thus they should be better identified from a brief presentation than items in the other two conditions. The results in Table 24.1 show exactly this pattern. The general point is that Jacoby (1983) and others (e.g., Kolers & Roediger, 1984; Morris et al., 1977), explain interactions among study and test conditions by appealing to the kinds of processing required during study and test manipulations. However, a very different form of explanation has been used to account for the next pattern of data to be described.

Another example of dissociation between tasks requiring different forms of remembering was reported by Tulving, Schacter, and Stark (1982). They presented subjects with 96 words for study and then tested them both 1 hour and 1 week later on two different tests. One was a standard yes/no recognition test and the other was a modification of a word fragment completion test that had been used previously in testing amnesics (Warrington & Weiskrantz, 1970, Woods & Piercy, 1974). In the Tulving et al. (1982) version, subjects were given word frames with letters omitted and asked to complete the word. To perform as a subject in the unprimed condition, complete the following fragments. (Allow 20 seconds each): _EX_ _ NT, _FL_ _ _ _ EL, _ REV_ _ CE,
Tulving et al.’s (1982) findings showed independence between recognition and fragment completion performance in two different ways. First, over the week delay, recognition performance dropped markedly, but priming in fragment completion (i.e., the benefit in completing fragments for studied relative to nonstudied words) did not decline at all, as shown in Fig. 24.1. (The completion rate for nonstudied words was .31 so performance in the fragment completion test was considerably enhanced by the study of the words.) Second, when recognition was tested within a session prior to fragment completion, fragment completion performance was stochastically independent of recognition performance. That is, subjects completed the fragments as well when they had previously judged a recognition item to be old as when they had judged it to be new. Tulving et al. (1982) interpreted this pattern of results as generally supporting the notion of separate memory systems (episodic and semantic), although they acknowledged some difficulties with this dichotomy and thought the results might implicate a third system, too. The important point is that dissociations among measures of retention are explained by appeal to different memory systems (see Tulving, 1983, Chapter 4, for description of the logic for postulating memory systems).

AIMS OF THE PRESENT RESEARCH

The foregoing research has indicated a consistent pattern showing dissociations among measures of retention. Independent variables often affect conscious recollection in one way, but exert no effect or even an opposite effect on other (transfer) measures that assess retention indirectly. The proper account of such dissociations among measures of memory is, to our minds, a central task facing cognitive psychologists today.

We turn first to the basic issues motivating the experiments. The major empirical aim of our studies is an examination of the effect of various surface features of information on its representation in memory and performance on

---

3 The words completing the fragments are SEXTANT, FLANNEL, CREVICE, AGNOSTIC, BOYHOOD, SWAHILI, INFERNO, BANDANNA, HYACINTH, DISSERTATION.

4 Episodic memory refers to the system that is thought to underlie experience for personal memories in which one must recollect the time and place of occurrence for an event to be remembered. Semantic memory is the system thought to be responsible for general knowledge in which one need not recall the time or place of occurrence to perform accurately. Tulving (1972, 1983, 1985b) spells out the distinction in greater detail. Within the context of Tulving et al.’s (1982) experiment, recognition memory taps the episodic memory system, since subjects must decide whether or not the test words belong to the studied list. Word fragment completion is a semantic memory task because subjects need not recollect previous occurrence of a word in the list to perform accurately.
FIGURE 24.1. Results of Tulving, Schacter, and Stark's (1982) experiment. Recognition dropped sharply over the retention interval, but fragment completion performance did not. However, word fragment completion was not simply insensitive, because studied words were completed better than nonstudied words. (Reproduced with permission.)

various memory tests. Fifteen years ago researchers had reached the conclusion that various media by which information was presented played little role in its long-term memory representation. Factors such as modality of presentation (auditory or visual), the typeface in which information was presented, or even the syntax (say, active or passive constructions) were believed to form no part of the trace in long term retention. Important studies by Sachs (1967), Bransford and Franks (1972), Kintsch and Monk (1972) and others bolstered this point, for these researchers found little effect of various surface characteristics on long-term recall or recognition. These studies led to the conclusion, which is still
prevalent in some quarters today, that incoming information is quickly coded into abstract representations (schemas, scripts, deep structures, logogens, or networks of associative connections). The medium seemed to play no role in the message coded in long-term memory.

These claims are, of course, challenged by many more recent findings, particularly those of Paul Kolers and his associates (e.g., Kolers, 1975, 1978; Kolers & Ostry, 1974; see Kolers & Roediger, 1984, for a review). Various surface features, such as modality of presentation or orientation of typeface (normal or rotated), are shown to have important consequences for performance on retention tests. One important difference between studies that have found no effect of form of presentation and those that have shown an effect is the dependent measures used. Typically, researchers showing that mode of presentation plays little role have measured recall or recognition, whereas Kolers’ studies have often measured speed of reading or other measures less dependent on conscious recollection (however, Kolers & Ostry, 1974, reported effects on recognition memory, too). It seems possible, if not likely, that the type of test is critical as to whether effects of surface form are revealed in performance. Jacoby’s (1983) studies reviewed previously are consistent with this claim.

In the experiments, we presented subjects with a list of items in a list and then tested their memories in various ways. We varied the modality (auditory or visual), typography (typed or handwritten, and upper- or lower case), language for bilinguals (Spanish or English), and form of referent (word or picture). The tests employed were free recall, yes/no recognition, and word fragment completion, although not all these tests were used in all experiments. Of primary interest is how these variables affect word fragment completion performance, a task that does not require conscious recollection. We show, as have Kolers and Jacoby, remarkable specificity in forms of presentation on certain retention tests. To anticipate, variables that have large effects on free recall exert no effect or even opposite effects on word fragment completion.

The general theoretical framework within which our experiments were designed was that of transfer appropriate processing (Morris, et al., 1977), in which improvements in test performance are expected to the extent to which the types of processing required on the test were acquired during the study episode. To state the same ideas in terms of processing operations (Kolers & Roediger, 1984), performance will benefit to the extent that procedures used in an acquisition phase are reinstated during a test phase. We later describe why we prefer these accounts of dissociations to others imputing various memory systems (e.g., Tulving, 1983).

**VARIATION IN MODALITY AND TYPOGRAPHY**

The effects of modality of presentation on recall have been examined in many experiments. The generally accepted conclusion, at least until lately, has been
that mode of presentation has an effect only on recently presented information (Crowder, 1976, Chapter 3). The conclusion aptly summarizing dozens of experiments on free and serial recall is that performance on the last few times is better if the information is presented auditorily rather than visually, but that very little or no effect of modality can be determined in long-term recall for information not recently presented (but see Gardiner & Gregg, 1979, and Glenberg, this volume, for a curious exception). Under conditions employed in our experiments to be described, modality has no effect on free recall (Blaxton, 1985).

To our knowledge, no studies of the effects of typography on free or serial recall exist, but data collected by Blaxton (1985) indicate that typography (uppercase elite type versus lowercase italic) has little effect in free recall. The case of modality and typography effects on single item recognition is somewhat different. Kirsner (1974) reported small effects of modality and typography on recognition, such that subjects performed better if the test form matched the surface form in these cases. Kolers, of course, has shown impressive effects of the orientation of print on the recognition of sentences (e.g., Kolers & Ostry, 1974).

Jacoby and Dallas (1981, Experiment 6) compared the effects of modality of presentation on yes/no recognition of visually presented test items and on their identification from brief displays. They reported a sharp dissociation, with modality having no effect on recognition but a large effect on perceptual identification. In the latter measure, performance was much improved if the item had been presented visually, but no benefit occurred for items presented auditorily (relative to nonstudied items). Winnick and Daniel (1970) and Clarke and Morton (1983, Experiments 1 and 2) have reported similar findings. This pattern can be accounted for by assuming that perceptual identification is a data-driven task (and thus should be highly sensitive to the way “data” are presented at study), whereas recognition is largely a conceptually driven task and is thus less affected by the form of presentation. However, according to Jacoby (1983), recognition also involves a data-driven component. Subjects may judge an item to be familiar when it is rapidly processed—when it seems to “jump off the page” (see Mandler, 1980 for a similar assumption). Thus, to the extent that recognition judgments are affected by such perceptual fluency factors, one can account for the effects of surface features such as found by Kirsner (1974) and Kolers and Ostry (1974) in recognition. Johnston, Dark, and Jacoby (1985) provide evidence for the operation of a data-driven component of recognition judgments under some conditions.

In the first of our experiments described here, (described in part in Roediger & Blaxton (1987) as Experiment 1), we presented subjects with 96 words, half visually and half auditorily. The items were presented under one of four conditions, in blocked fashion. Half of the 48 visually presented items were typed on an IBM Selectric typewriter in lowercase letters; the other 24 visually presented items were printed by hand in uppercase letters (all visually presented items were
shown to subjects via a slide projector). Twenty-four other items were presented auditorily; 24 more were also presented auditorily, but subjects were told to form an image of the word as it would appear typed (subjects were given prior experience with the typeface). Across subjects, all items appeared equally often in each study condition and were presented at a 5-second rate.

These various study conditions were manipulated within subjects, but type of test was a between-subjects factor. One set of subjects received a standard yes/no recognition test in which the 96 old items were randomly intermixed with 96 new items on a sheet with instructions to circle items they recognized as being old. Each test typography was used for half the items in each study condition. Subjects who received the word fragment completion test received exactly the same sort of form as did recognition subjects (i.e., half old and half new items, etc.), but now only fragments of the words were presented and the subjects’ task was to fill in the missing letters to complete the word correctly. (Examples of our materials were provided earlier.) The design appears in Fig. 24.2.

The word fragment completion test results are presented in Table 24.2. Note first that, relative to the nonstudied completion rate, items presented under all study conditions primed their later completions. Thus, unlike Jacoby and Dallas’s (1981) finding with perceptual identification (their Experiment 6), we obtained cross-modal priming in fragment completion, in that completion rates averaged 0.43 in the auditory condition and only 0.27 in the nonstudied condition. However, priming from visual presentations was greater than for auditory presentations (0.52 versus 0.43), showing that same-mode priming exceeded cross-modal priming. In addition, within the visual mode, a slight but significant effect of typography was obtained. When the typography at test matched that of the study episode, performance was better than when the two mismatched (0.55 and 0.50 completions, respectively; most of this effect is due to superior performance in the Handprinted–Handprinted case). Finally, when subjects were presented words auditorily but told to image what the word would look like typed, fragment completion performance improved about 5% relative to the case of auditory presentation, (0.48 to 0.43), but this increase was not specific to the mode of test (i.e., the increase also appeared when subjects were tested with the handprinted fragments, contrary to expectation). This last finding seems to argue that subjects encoded more information when instructed to image the word, but that this effect was not specific to the typeface imagined. Nonetheless, it is interesting that when subjects were instructed to image the word typed, word fragment completion performance was equivalent to that of items actually presented typed (see Jacoby & Witherspoon, 1982, for a similar result).

Recognition results for the experiment are shown in Table 24.3, where the trends found in fragment completion are largely absent. That is, visual presentation was not reliably superior to auditory presentation, nor was there a reliable effect of typeface. Whereas visual presentation produced the best performance in fragment completion, auditory presentation with instructions to image the words
Roediger & Blaxton (1987)
Design of Experiment 1

<table>
<thead>
<tr>
<th>Study Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>96 Words</td>
</tr>
<tr>
<td>48 Visual</td>
</tr>
<tr>
<td>48 Auditory</td>
</tr>
<tr>
<td>24 Handprinted (Uppercase)</td>
</tr>
<tr>
<td>24 Typed (Lowercase)</td>
</tr>
<tr>
<td>24 No Instruction</td>
</tr>
<tr>
<td>24 Imagine Typed in Lowercase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment completion:</td>
</tr>
<tr>
<td>192 Fragments</td>
</tr>
<tr>
<td>96 Studied</td>
</tr>
<tr>
<td>48 Handprinted</td>
</tr>
<tr>
<td>48 Typed</td>
</tr>
<tr>
<td>96 Nonstudied</td>
</tr>
<tr>
<td>48 Handprinted</td>
</tr>
<tr>
<td>48 Typed</td>
</tr>
<tr>
<td>Recognition memory:</td>
</tr>
<tr>
<td>192 Words</td>
</tr>
<tr>
<td>96 Studied</td>
</tr>
<tr>
<td>48 Handprinted</td>
</tr>
<tr>
<td>48 Typed</td>
</tr>
<tr>
<td>96 Nonstudied</td>
</tr>
<tr>
<td>48 Handprinted</td>
</tr>
<tr>
<td>48 Typed</td>
</tr>
</tbody>
</table>

FIGURE 24.2. The design of Roediger and Blaxton's (1987) Experiment 1. The study phase manipulation is represented at the top, with the test phase manipulations below.

| TABLE 24.2 |
| Proportion of Fragments Completed in Each Condition of Experiment 1 |

<table>
<thead>
<tr>
<th>Study Condition</th>
<th>Visual</th>
<th>Auditory</th>
<th>Nonstudied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Condition</td>
<td>Handprinted</td>
<td>Typed</td>
<td>No Instruction</td>
</tr>
<tr>
<td>Handprinted</td>
<td>0.59</td>
<td>0.48</td>
<td>0.44</td>
</tr>
<tr>
<td>Typed</td>
<td>0.52</td>
<td>0.51</td>
<td>0.42</td>
</tr>
</tbody>
</table>
yielded best recognition. A slight tendency existed, within the visual presentation conditions, for compatible typographies to produce better performance than mixed typographies, but this trend was not statistically reliable. Although some dependency may exist across conditions between the recognition memory and fragment completion tests, its magnitude seems small.

We performed a second experiment on the issue of effects of typography and modality on fragment completion, this time including delay of recall. Subjects were tested either a few minutes or one week after studying the list. In general, the main results of the first experiment were replicated at both testing intervals. Priming occurred in all study conditions, with visual presentation leading to greater priming that auditory presentation, and matched typographies at study and test producing better performance than mismatched typographies. Notably, we found significant priming on the word fragment completion task after a 1-week retention interval, although performance declined in all conditions during the delay. Subjects completed 10% fewer fragments in the primed conditions after the week delay. Tulving et al. (1982) found no such loss in priming over a week, but their “initial” test occurred an hour after presentation whereas ours was given shortly after study. This procedural difference may account for the discrepancy.

In sum, the results of these experiments indicate that fragment completion is highly sensitive to the correspondence between features of the study and test presentations for both modality and typography. Recognition is less sensitive to typography in our experiments and, if anything, the effect of modality on recognition was opposite that for fragment completion (i.e., a slight tendency existed for auditory presentations to exceed visual in recognition).

Our results agree with those of Graf, Shimamura, and Squire (1985, Experiment 1), who reported cross-modal priming for both amnesics and normals in a word stem completion task in which subjects received three letters of a word with instructions to provide the first completion that came to mind. Same-mode priming was also greater than cross-mode priming. Our results and those of Graf et al. (1984) show that priming can occur across modality boundaries, unlike results
obtained by others using perceptual identification (Jacoby & Dallas, 1981, Experiment 6) or threshold measures (Clarke & Morton, 1983; Winnick & Daniel, 1970). Our results on the effects of typography also provide tentative support for specificity of visual features in priming, unlike the results of Clarke and Morton (1983; see also Jackson & Morton, 1984).

THE EFFECTS OF LANGUAGE OF STUDY

The next experiment we discuss was conducted by Aydin Durgunoğlu in collaboration with the first author and tested how language of presentation affects performance in bilingual subjects (Durgunoğlu and Roediger, 1987). The subjects were Spanish–English bilinguals, mostly graduate students at Purdue whose native language was Spanish but who were also fluent in English. All subjects were presented 115 words to remember in six study conditions and then received one of three types of test: free recall, yes/no recognition, or word fragment completion. Only some of the study conditions need be described for present purposes; these are illustrated in Table 24.4. The study conditions were varied within subjects, while the type of retention test was a between-subjects factor. Sets of words were rotated through the study conditions across subjects and were presented individually at an 8-second rate.

The study conditions shown in Table 24.4 indicate that during the 8-second study period, subjects either read the words (a) twice in Spanish or (b) twice in English, or (c) once in each language. The first presentation appeared on a screen via a slide projector and the second was given on a sheet before the subject. Items in these study conditions were presented in blocks, prefaced by the appropriate instructions. One set of items was not presented but was used to assess base line

<table>
<thead>
<tr>
<th>Study Condition</th>
<th>Type of Test</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Free Recall</td>
<td>Yes/No Recognition</td>
<td>Word Fragment Completion</td>
</tr>
<tr>
<td>Read Spanish caballo, caballo</td>
<td>0.23</td>
<td>0.62</td>
<td>0.44</td>
</tr>
<tr>
<td>Read English horse, horse</td>
<td>0.17</td>
<td>0.79</td>
<td>0.66</td>
</tr>
<tr>
<td>Read Spanish &amp; English caballo, horse</td>
<td>0.39</td>
<td>0.95</td>
<td>0.68</td>
</tr>
<tr>
<td>Nonstudied</td>
<td>—</td>
<td>0.05</td>
<td>0.42</td>
</tr>
</tbody>
</table>

TABLE 24.4
Results of Durgunoğlu and Roediger (1987)
(Reproduced with permission.)
performance on the fragment completion and recognition tests. On these latter tests, subjects received words (for recognition) or word frames (for fragment completion). Recognition subjects were told to circle words they recognized as having occurred in the list and were explicitly told to circle words referring to concepts that may have occurred in the alternate language. Fragment completion subjects were simply told to write the first word that came to mind to complete the fragment. All items in the fragment completion and recognition tests were given in English, a fact impressed on the subjects through instructions just before the test. For the free-recall test, subjects were instructed to recall as many words as possible from the study list in whichever language they came to mind.

The results are presented in Table 24.4. First, consider free-recall performance. Items presented in Spanish were somewhat better recalled than those given in English, probably reflecting the fact that our subjects were mostly Spanish-dominant bilinguals (however, the difference was of borderline significance). Interestingly, recall was better when subjects read the word in both languages during study rather than twice in the same language, even when this was Spanish. This may reflect an effect of encoding variability (Madigan, 1969), and replicates a finding of Glanzer and Duarte (1971) with massed presentation. In Jacoby’s (1983) terms, free recall is a conceptually driven task because no “data” are given in the task to guide performance. In line with his findings, the condition that encouraged coding of conceptual information (variable encoding) produced better performance than the two conditions that simply required subjects to read information in one language.

Turning to the word fragment completion results on the far right of Table 24.4, a very different pattern emerges. Relative to the completion rate for new items (0.42), studying the word twice in Spanish produced no reliable priming (0.44), (Watkins & Peynircioğlu, 1983, Experiment 2 reported similar findings). However, reading the word in English, or reading it in both Spanish and English, produced sizable priming of roughly equal magnitude. Thus, fragment completion acts as a data-driven task. Subjects use the fragments as “data” to construct words. If similar data had been presented at study (i.e., words in English), performance was facilitated.

According to Jacoby (1983), recognition involves a mixture of both conceptual and data-driven processing. To use Mandler’s (1980) terms, there is both activation or integration of the surface features of the item, and the elaboration of concepts. Thus, recognition should show a mixture of the features of free recall and word fragment completion, depending on whether study conditions emphasized processing of surface features or the elaboration of concepts. This is true in our experiment on bilinguals. Several lines of evidence for conceptual aspects of recognition are present. First, recognition shows better than chance responding in all conditions (unlike fragment completion), presumably reflecting conceptually driven processing. Second, subjects recognized words better after the elaborative processing involved in the varied encoding condition (presentation in
both languages, 0.95) than when the items were presented twice in the same language (0.62 and 0.79). However, evidence also exists for a data-driven component to recognition. In recognition (unlike free recall), presentation of the items in English (0.79) led to better performance than their presentation in Spanish (0.62). This presumably reflects the fact that the test was all in English and thus with English presentation there was a greater match between the data at study and test.

In sum, a reasonable story can be made from results of the present experiment indicating that free recall is conceptually driven, fragment completion is largely data-driven, and recognition memory involves a mixture of the two types of processing. The data are thus in line with the proposals of Jacoby (1983), as well as related ideas of Graf and Mandler (1984).

PRIMING BY PICTURES AND BY WORDS

The next experiment to be discussed under the general rubric of effects of surface aspects of media on retention involves a comparison of memory for pictures and words. The general finding from a huge body of literature is that pictures are recalled and recognized better than their labels (Paivio, 1969, 1971). Free recall shows a substantial picture advantage and so does recognition. In free recall, the picture advantage occurs despite the fact that subjects are actually recalling words (labels of the pictures) after studying pictures, a factor which might be expected to favor word recall. Interestingly, the advantage of pictures in recognition also occurs even when the test items are presented as words. That is, if the word horse were presented during the test, subjects would recognize it better if they had previously studied a drawing of a horse rather than the word itself. Even though the data at test better match those studied in the word–word case then in the picture–word case, recognition is better in the latter (see Madigan, 1983). The usual interpretation of these findings is that pictures are encoded into a richer or more elaborate representation in memory and therefore are better recognized and recalled than are words (e.g., Nelson, Reed, & McEvoy, 1977; Paivio, 1971).

The question motivating the present research was whether or not the usual picture superiority effect could be eliminated or even reversed when the word fragment completion test was used to assess retention. Assuming that fragment completion is a data-driven task, then medium of presentation should play a larger role in determining fragment completion than it does in recognition or recall. The experiment described here is the first in a series conducted by Mary Susan Weldon in collaboration with the first author (Weldon & Roediger, 1987). Thirty-three items were selected from the Snodgrass and Vanderwart (1980) picture norms and prepared in two forms, either as simple line drawings or as the labels for the drawings. Twenty-two items were presented to subjects, half as
TABLE 24.5
Results from Weldon and Roediger (1987)
(Reproduced with permission.)

<table>
<thead>
<tr>
<th>Study</th>
<th>Free Recall</th>
<th>Word Fragment Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drawing</td>
<td>0.35</td>
<td>0.45</td>
</tr>
<tr>
<td>Word</td>
<td>0.28</td>
<td>0.64</td>
</tr>
<tr>
<td>New</td>
<td>—</td>
<td>0.38</td>
</tr>
</tbody>
</table>

drawings and half as words, in the midst of a number of buffer items. All
subjects received list presentation under the same conditions, but half were given
a free-recall test and the other half received a word fragment completion test. In
the fragment completion test, subjects completed 33 fragments representing 11
items studied as drawings, 11 as words, and 11 that were not studied. Items were
counterbalanced across the study conditions over subjects. The primary results
are shown in Table 24.5, in which the proportion of items recalled and completed
in each of the conditions is given. As usual, pictures were better recalled than
words (0.35 to 0.28). However, this pattern was reversed in the word fragment
completion test. Relative to the base line, pictures produced only a seven percent
priming effect, but study of words led to a 27% priming effect. The interpreta-
tion, which should be familiar now, is that the word fragment completion test,
being data-driven, is facilitated by a match between mental operations engen-
dered by processing the data in the study and test phases. Free recall, which
depends on elaborative processing, reveals the alleged “richer” encoding of
drawings than words.

The reversal of the usual picture superiority effect with word fragment com-
pletion as the measure of retention is not due simply to a perceptual match
between the data (cues) given at test and those studied earlier, but must also be
due to the type of retrieval operations subjects perform on the data provided
during testing. The reason for this assertion is a comparison between our findings
and those described earlier employing recognition of words (Madigan, 1983).
Surely the perceptual match between the word studied and the test cue is greater
when recognition memory is tested (e.g., *birdhouse* at study and test), than when
the subjects’ task is word fragment completion (*birdhouse* at study and b___d
h___e at test). In recognition, pictures are recognized better than words,
but in fragment completion words are identified better than pictures. Thus the
task demands, and not simply the completeness of the cues, helps determine
performance. Presumably, on a recognition test, subjects call conceptual opera-
tions into play that are largely absent on the word fragment completion task. We
say “largely” and not “completely” absent, for we did find modest but reliable
priming from pictures on word fragment completion (see Table 24.5). Of course, this effect may be explained by covert labeling of pictures during study, but in later experiments (Weldon and Roediger, 1987, Experiments 2 and 3) we found no change in primed fragment completion levels from pictures when we varied instructions designed to encourage or discourage labeling. Thus the priming effect of pictures on word fragment completion may be conceptually mediated, a point to which we will return.

REPRISE: MEDIA AND RETENTION

A few years ago the commonplace assumption was that the medium of presentation had little to do in most cases with the underlying message retained in long-term memory. The present results add to those of others (Jacoby & Dallas, 1981; Kolers, 1975; Kolers & Ostry, 1974) in showing that this conclusion is false. Although some tests are insensitive to changes in surface representation, others reveal large effects. Typography, modality, and even language of presentation for fluent bilinguals are all variables that have no effect on long-term free recall in most situations, and little effect on recognition. However, all these variables yield large effects on word fragment completion as seen in the present experiments, and they also affect other measures of implicit retention such as perceptual identification (Jacoby & Dallas, 1981) visual duration thresholds (Winnick & Daniel, 1970), word stem completion (Graf et al., 1985), and reading speed (Kolers, 1975).

These recent results also add new information about word fragment completion as a measure of retention. Presented in Fig. 24.3 is a graph showing the amount of priming obtained from studying items in various forms. Constructing this graph takes a certain amount of daring because different sets of items were used across the experiments reviewed here. However, each of these experiments (and one other, not yet described) includes a condition involving repetition priming (same mode and typography at study and test). The amount of repetition priming across five experiments was 0.24, 0.26, 0.27, 0.28, and 0.31 for an average of about 0.27. These values seemed stable enough to warrant the construction of Fig. 24.3.

The data in Fig. 24.3 provide a reasonable ordering of influences, convincing us that word fragment completion does indeed represent a "data-driven" test of retention, in Jacoby's (1983) terms. The magnitude of the priming effect generally accords with expectations based on similarities between data processed at study and the fragment provided at test. Performance is best when presentation is visual and the typography matches that used at test (0.27), and is next best with visual presentation and mismatched typography (0.23) (this effect of typography was larger in Experiment 1 of Roediger & Blaxton, when materials were held constant; the diminution here occurs from averaging across experiments to obtain
the estimate of repetition priming). When words were presented auditorily, less priming occurred (0.16), although this effect is accentuated slightly if subjects are told to image the appearance of the word. Performance was worst when the presentation of the item occurred in a very different format or medium from the fragment—pictures produced a 0.07 priming effect, forming an image to a word presented in a different language led to a 0.06 priming effect, and presenting the item in an alternate language produced essentially no priming effect (0.02).

The failure to find substantial priming effects in these latter three cases cannot be attributed to subjects' not encoding the list items, since free-recall tests actually showed better performance in these three conditions than in the appropriate control conditions. The conclusion is that fragment completion does not rely on elaborative or conceptually driven processing, but instead on data-driven processing.
The three conditions at the far right of Fig. 24.3 come from an experiment conducted in collaboration with Valerie Ludwick and Bradford Challis that has not yet been described. Briefly, subjects studied a list of words and then received a word fragment completion test similar to those described earlier. The words in the list bore various relations to the items represented in the word fragments. Besides repetition priming (study horse and tested with h__r__e, for example), studied items could bear other relations to the fragments such as being synonyms (stallion), associates (saddle), or category coordinates (cow). As can be seen in the figure, these types of studied items produce very little priming on the word fragment completion test. Synonyms and associates produced about a three percent effect, relative to the nonstudied base line, and coordinates produced no priming at all. Again, the standard word fragment completion test seems largely data-driven, because when items are given that are meaningfully related to the fragment but whose appearance is quite different, little or no priming occurs.

THEORETICAL ALTERNATIVES

Throughout our description of the above research, we have employed a procedural view (Kolers & Roediger, 1984) in which performance on a retention or transfer test is assumed to be a function of the similarity of processing operations engendered between study and test, with further assumptions added about the types of processing required by different tests (Jacoby, 1983; Mandler, 1980). We will defend this view, but will first consider our results as they might be interpreted by other theories.

Strength Theory

At the considerable risk of flogging a long-deceased caballo, let us consider the implications of these results for strength theory. A simple strength theory would maintain that experiences leave memory traces as their mental residue and that these vary in strength along a single dimension. Different levels of performance reflect varying strengths in these traces. The assumption that remembering is determined only by a trace dependent factor, such as strength, makes the obvious prediction that the ordering of performance in various study conditions should generalize across memory tests. That is, if Trace A is stronger than Trace B on a free-recall test, then this ordering should be preserved on all other tests. Obviously, the present results and many others falsify this hypothesis. Any study showing opposing effects on memory tasks as a function of independent variables, as in the Weldon and Roediger (1987) experiment described earlier, can be so considered. Other examples include Balota and Neely (1980), Fisher and Craik (1977), and Morris et al. (1977).
Logogens and Nodes

Priming effects in tasks such as lexical decision or naming latency are often attributed to the activation of a logogen or node that is effected by external experiences. Such hypothetical entities are usually postulated to be relatively abstract, and do not incorporate surface features of the medium by which the information is conveyed. For example, according to Morton’s (1969) theory, the same logogen would be aroused if one saw the word *horse*, or heard someone say it.

Many recent results challenge the logogens view, including those of Morton (1979). These experiments replicate and extend the earlier work of Winnick and Daniel (1970), showing that prior visual presentation of a word produced greater facilitation in reducing visual duration thresholds than did auditory presentation. Morton (1979) postulated separate visual and auditory logogen systems to account for the results. Experiments by Jacoby and Dallas (1981) and Jacoby (1983) raise similar problems and even extend them, for Jacoby and Dallas (1981) found substantial priming in a perceptual identification task after a 2-day retention interval. The activation of logogens is usually thought to be short-lived, on the order of seconds, and thus priming after a 2-day interval stretches the concept beyond recognition.

The present results also discredit the logogen/node account of repetition priming, although in fairness the idea was never put forward to encompass situations such as those investigated here. In order to account for priming in word fragment completion, the logogen account would have to postulate (a) activation lasting up to a week; (b) different logogen systems not only for modality of presentation, but also for different typographies, for presentations in alternate languages, and for presentation of an “item” as a word or picture. The attractiveness of the idea of abstract logogens as an explanatory device is reduced if a new logogen system must be invented for every variable that affects priming, and if one of the concept’s basic properties—activation that decays rapidly—is removed. Simpler alternatives that would provide a more parsimonious account are therefore preferred.

Memory Systems

The results reported above reveal several dissociations between independent variables and performance on the various memory tests. Tulving (1983, Chapter 5) has interpreted similar dissociations as evidence for separate memory systems, episodic and semantic. The former is believed responsible for recollection of personal experiences requiring retrieval of time and place, whereas the latter underlies permanent knowledge. Most dissociations in the evidence reviewed by Tulving took the form of independent variables affecting an episodic memory task (such as recognition), but having no effect on a semantic memory test (such
as fragment completion or perceptual identification). For example, Jacoby and Dallas (1981) manipulated the “level of processing” of presented words by having subjects attend to their appearance, sound, or meaning and then provided either a perceptual identification or a yes/no recognition test. They found that level of processing had the usual large effect on recognition, but no effect on perceptual identification. Tulving (1983) argued that such results support the distinction between episodic and semantic memory, because a variable is shown to affect a task tapping one type of memory (episodic recognition) but not the other (perceptual identification).

Our own experiments reviewed earlier reveal several dissociations that can be similarly interpreted. Modality and typology were shown to affect word fragment completion, but not recognition. This outcome is different from virtually all those reviewed by Tulving (1983) in showing variables that exert greater effects on a semantic memory task than on an episodic memory task. Two other variables studied here produced opposing effects on recall and fragment completion, a more powerful dissociation similar to that of Jacoby (1983). For bilinguals, words presented in Spanish were better recalled than words given in English, but the opposite was true in word fragment completion (with English fragments). Similarly, pictures were free recalled better than words, but words primed fragment completion more than pictures. Again, these interactions can be interpreted as evidence for different memory systems, with one system assumed to underlie free recall (episodic memory) and the other to underlie word fragment completion (semantic or, perhaps, procedural memory; see Tulving et al., 1982).

The use of dissociations to support the episodic/semantic distinction has been questioned on several occasions (Hintzman, 1984; McKoon, Ratcliff, & Dell, 1986; Roediger, 1984). One argument is that the theory is not specific enough to predict the form of the interaction in a given experiment. The general form of Tulving’s argument (1983, Chapters 4 and 5) can be paraphrased as “my theory predicts interactions between study and test conditions; interactions are found; therefore the theory is supported.” Reasons are rarely given as to why some variable should affect an episodic but not a semantic memory task, or vice versa. In fairness, in an enterprise so new as trying to distinguish among various memory systems, perhaps Tulving’s logic of simply looking for interactions is the best possible without further knowledge. However, it is only persuasive if some other framework or theory cannot provide a better account of these interactions. As we have argued throughout, we believe an alternative is available, a matter to which we now turn.

Transfer Appropriate Procedures

We have argued against the ideas of separate memory traces with varying strengths as an explanation of our results; we argue too against the ideas of separate logogens or nodes that are activated, and against proliferating memory
systems. Instead we opt for a more process-oriented view, emphasizing the mental operations people perform in accomplishing tasks and the information these require. The ideas represent an amalgamation of those from Kolers and Roediger (1984), Jacoby, (1983), and Morris et al. (1977; see also Bransford, Franks, Morris, and Stein, 1979). In addition, they are similar in spirit to the encoding specificity hypothesis (Flexser & Tulving, 1978; Tulving and Thomson, 1973), although the latter emphasizes mental contents more than processes. Here we first state the ideas rather abstractly, and then will use them to interpret data from the experiments which have been described. Third, we describe data from another experiment that we believe supports this view over Tulving’s (1983) postulation of separate memory systems.

When one is exposed to information, the instructions given, the type of material, its expected future use, and individual proclivities combine to emphasize certain types of mental operations (Jenkins, 1979). To take familiar examples, subjects’ attention can be directed at various features of words, as in the ‘‘levels of processing’’ experiment; or subjects can be told to form mental images; or they can be led to expect recall, recognition, or some other type of test. The processing activities so engendered affect the type of information learned about an experience—what is coded—and that in turn affects how information can be expressed on some particular test. In general, the more similar the processing activities required by the test to the encoding activities, the better will be performance on the test. This restates the ideas of many others, (e.g., Bransford et al., 1979; Kolers, 1975; Kolers & Roediger, 1984; Tulving & Thomson, 1973), although in somewhat more general form.

We find particularly useful the additional idea, applied by Jacoby (1983) to remembering, of data-driven (bottom–up) and conceptually driven (top–down) processing. Some study and test conditions will emphasize attention to presented information (the data), whereas others will emphasize concepts—elaborations, associations, etc. (see also Mandler, 1980, and Graf & Mandler, 1984). Examples of these ideas as applied to study activities would be judgments of appearance versus judgments of meaning in the levels of processing experiment; reading versus generating in the generation effect experiment; and rote repetition of words versus forming images of their referents in experiments on mental imagery (e.g., Bower, 1972). In each case, the first activity emphasizes data more than concepts and the second does the reverse.

Turning to the various memory tests, free recall is the paradigmatic ‘‘conceptually driven’’ test, for no data are presented as cues. Rather, subjects must ‘‘drive’’ their performance by using mental concepts. Other tests such as perceptual identification and word fragment completion are more data-driven. That is, subjects are given data at test and must process it as rapidly as possible in performing some task. When used as tests of retention, the data in these tasks are provided in no particular context, and thus conceptually driven processing is difficult (especially in perceptual identification).
The distinction between data-driven and conceptually driven processing should not be considered a strict dichotomy, we believe, but rather as endpoints on a continuum. For example, a word fragment completion test could be made more conceptually driven if subjects were given associative cues (DAGGER — A ___ A ___ IN). Similarly, recall is made more data-driven if one provides cues that resemble the targets in terms of their appearance (Blaxton, 1985) or sound (Fisher & Craik, 1977). Recognition memory tests are argued to be affected both by data-driven and conceptually driven components (Jacoby, 1983), or by activation of features and elaboration of meaning (Mandler, 1980). However, in most standard yes/no recognition memory tests, with their leisurely pace, conceptually driven processing is probably more important (see Johnston et al., 1985).

Having provided these rather abstract distinctions and assumptions, how might they apply to our data? In these experiments we manipulated the medium of presentation—modality (auditory or visual), typography (uppercase and hand-printed, lower case and typed), language (Spanish or English), and form of representation (drawing or word). In general, our results fit well with the ideas outlined here. Variation in modality and typography affected the data-driven word fragment completion test, but not the recognition test that is believed to have a greater conceptually driven component. (Others have found slight effects of these variables on single word recognition, e.g., Kirsner, 1974.) However, Blaxton (1985) found no effect of typography on free recall, and a large literature on modality effects reveals no effect on free recall in long-term memory under conditions similar to those employed here (Crowder, 1976; Murdock & Walker, 1969). Thus the effects of typography and modality fall in line with what one would expect—largest effects on data-driven tasks, small effects on recognition with its small data-driven component, and no effect on free recall.

Language of presentation (Spanish or English) and form of representation (drawing or word) provide more interesting manipulations of media, because in both cases one can make a reasonable assumption that one mode of presentation produces a richer, more elaborate encoding than does the other. For our Spanish-dominant bilinguals, words in Spanish may be expected to be more richly encoded than words in English; similarly, drawings are assumed by many to be more elaborately encoded than their verbal labels (e.g., Paivio, 1969). In line with this assumption, the relevant experiments reported here showed slightly greater free recall for Spanish than English words by our bilinguals, and for pictures than words. However, on the word fragment completion test, performance reversed in both cases. In this case, bilinguals completed the words better than had been presented in English than those presented in Spanish (the fragments were presented in English). For our other subjects, presenting items as words produced greater priming in word fragment completion than presenting them as pictures. In both cases, these outcomes were predicted by the theory emphasizing transfer appropriate processing, since the data-driven fragment
completion test should benefit most from conditions in which similar data are processed during study, viz. English words.

A COMPARISON OF THEORIES

We have shown that the transfer appropriate procedures approach can account well for the dissociations in our data. But so can theories that ascribe these dissociations to different memory systems, as we have outlined. How can we decide which approach is best? Or are both approaches so vague that no distinctive, testable consequences can be derived from them?

We believe that the transfer appropriate processing approach is clearly better, and that testable differences can be derived. Consider first some data from our own experiments. In the Durgunoglu and Roediger (1987) experiment, some items were studied only in English and others only in Spanish, among the other conditions. In free recall, Spanish words were slightly better remembered than English words, but in recognition (which gave test words only in English) words that had been presented in English were better recognized. How can the memory systems approach explain this finding? Both recognition and recall are episodic memory tests, so the same pattern of results is to be expected in both tasks. Presumably, dissociations should not occur within tasks tapping the same system. On the other hand, the view emphasizing transfer appropriate procedures has no trouble with this finding. Recognition, unlike free recall, has a data-driven component and thus on a test in which English words are given, having studied English words should be better than having studied words in another language.

A critic of our logic might complain, somewhat justifiably, that we stacked the deck in this experiment by giving these test items in English. All Tulving (1983, 1985b) would need to do is add something like transfer appropriate processing to the multiple systems idea to account for the results. That is true enough, but such a course then raises the interesting question of why the notion of memory systems need be added at all to explain the results. The transfer appropriate procedures idea accounts for the data without requiring the addition of separate systems.

Also, the switch between Spanish and English in our experiment does not affect the form of the argument, since the same point can be made from other data. For example, if word frequency is manipulated in recognition and recall experiments, the consistent finding is that high-frequency words are better recalled than low-frequency words, but that low-frequency words are better recognized than high-frequency words (e.g., Balota & Neely, 1980). Once again, there is a dissociation within two episodic memory tasks, which is difficult to encompass in Tulving’s (1983) system. On the other hand, if we make the common assumption that high-frequency words are more richly or elaborately
encoded than low-frequency words (Anderson & Bower, 1972), but that low-frequency words are more distinctive and receive more data-driven processing, then the explanation from the transfer appropriate procedures view assumes its familiar form.

In brief, Tulving’s (1983) view distinguishing among various memory systems does not provide a straightforward account of dissociations within tasks believed to tap the same system. This problem usually does not arise in experiments showing dissociations between measures of retention (see Tulving, 1983, Chapter 5), because researchers have typically employed only two tests of retention—one episodic and one semantic—when examining dissociations (Roediger, 1984). For example, Jacoby and Dallas (1981) looked at how various independent variables affected perceptual identification and yes/no recognition, and Tulving et al. (1982) examined the effects of retention interval on word fragment completion and yes/no recognition. With only two measures in an experiment—one thought to tap semantic memory and one episodic—any interaction obtained can be taken as support for the distinction between memory systems. On the other hand, the failure to obtain an interaction is not necessarily evidence against the distinction, because no one would argue that all variables should affect the two systems differently.

In most experiments showing dissociations between memory tests reviewed by Tulving (1983, Chapter 5), a confounding exists between the system the task is to tap and the type of task, according to Jacoby’s (1983) typology. Shown in Fig. 24.4 is a 2×2 representation crossing memory systems (episodic and semantic) with task type (data-driven and conceptually driven). Most experiments to date have involved only two cells—the episodic memory tasks have been con-

<table>
<thead>
<tr>
<th>Type of Task</th>
<th>Episodic</th>
<th>Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data-Driven</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptually-Driven</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 24.4.** Crossing of two memory systems with types of processing. Most previous experiments have involved a comparison of conceptually driven episodic memory tests with data-driven semantic memory tests, the lower left and upper right cells.
ceptually driven (recognition, recall) and the semantic memory tasks have been data-driven (perceptual identification, word fragment completion).

The design of experiments needed to distinguish the memory systems approach to explaining dissociations from the transfer appropriate procedures view is to use all four tasks, as suggested by Fig. 24.4. That is, tests involving episodic and semantic memory should be varied orthogonally to the type of processing the task requires. Assume that two tasks can be reasonably classified as episodic and two as semantic; also that two of these can be classified as data-driven and two as conceptually driven. If these dimensions are uncorrelated, it should be possible to vary study conditions and assess performance on the four different tasks to determine whether the results better conform to an interpretation in terms of the memory system that each task involves or the type of processing each requires. Blaxton (1985) has used this general logic in a series of experiments designed to determine the adequacy of these two approaches to explaining dissociations. Here we will describe only one of her experiments that makes the general point. In addition, we will simplify our discussion of the design slightly to make our points more clearly.

Subjects in Blaxton’s (1985) Experiment 1 studied a list of words under three conditions analogous to those in Jacoby’s (1983) experiment. On some occasions subjects received items such as XXX–COPPER (no context); on others they received items in context, in the form of a semantic associate (tin–COPPER); and in still other cases they had to generate the word (TIN–C___）。

This manipulation was intended to affect the relative amounts of data-driven and conceptually driven processing during study, with data-driven processing decreasing across the no context, context, and generate conditions, but with conceptually driven processing increasing across these conditions. These conditions were varied within-subjects, in blocks of 21 items. Thus subjects studied 63 words altogether before receiving a test, the nature of which was not specified beforehand.

Four different tests were given, as can be seen in Fig. 24.5. Test type was manipulated between subjects. Free recall is considered a conceptually driven episodic memory test. The data-driven episodic memory test created by Blaxton (1985) was cued recall with words that looked like the target words (e.g., CHOPPER for COPPER.) Subjects taking this test were carefully instructed that they were to try to recall words that looked like the cue words and were told that the meaning of the cues was irrelevant to the task. The data-driven semantic memory task was word fragment completion, on which these subjects told simply to think of the first word possible to complete the word fragment. Half the fragments given were for studied words and half for nonstudied words, with the two sets counterbalanced across subjects. The conceptually driven semantic memory test was a test of general knowledge (Nelson & Narens, 1980), e.g., “What metal makes up ten percent of yellow gold?” For half the questions given
Memory System

Type of Task

<table>
<thead>
<tr>
<th>Episodic</th>
<th>Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphemic Cued Recall</td>
<td>Word Fragment Completion</td>
</tr>
<tr>
<td>Free Recall</td>
<td>General Knowledge Test</td>
</tr>
</tbody>
</table>

1. Graphemic Cued Recall: CHOPPER
2. Free Recall: No cues given
3. Word Fragment Completion: C __ P __ E __
4. General Knowledge Test: What makes up 10% of yellow gold?

FIGURE 24.5. Test conditions used in Blaxton (1985, Experiment 1).

on the test, the correct answer was a previously studied word and for half it was not. Thus, for the two semantic memory tests, a base line of nonstudied words was included to assess priming from prior study.

The results are presented in Table 24.6. If the memory systems approach to explaining dissociations between study and test conditions is accurate, then we should expect to see performance similar in the two episodic tasks (free recall and cued recall with graphemic cues), and different from that in the two semantic tasks (word fragment completion and accessing general knowledge). On the other hand, if the transfer appropriate processing view is more accurate, we should expect to see similar performance in the tasks emphasizing data-driven

<table>
<thead>
<tr>
<th>Study Condition</th>
<th>No Context</th>
<th>Context</th>
<th>Generate</th>
<th>Nonstudied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free recall</td>
<td>.19</td>
<td>.16</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>General knowledge</td>
<td>.33</td>
<td>.38</td>
<td>.50</td>
<td>.25</td>
</tr>
<tr>
<td>Graphemic cued recall</td>
<td>.45</td>
<td>.40</td>
<td>.34</td>
<td>.06</td>
</tr>
<tr>
<td>Word fragment completion</td>
<td>.75</td>
<td>.62</td>
<td>.46</td>
<td>.27</td>
</tr>
</tbody>
</table>

TABLE 24.6
Results of Blaxton (1985) Experiment 1
(Reproduced with permission.)
processing for their completion (word fragment completion and graphemic cued recall) and a different pattern in the tasks requiring conceptually driven processing (free recall and accessing general knowledge). As is apparent from Table 24.6, the data support the transfer appropriate processing view. As expected, generated items produced better performance than those read without context in both the free recall and general knowledge tests. These are both supposed to involve conceptually driven processing, although they tap different memory systems. However, for word fragment completion and graphemic cued recall (the data-driven tests), this pattern is reversed. Here, items studied in the No Context condition transferred better to the test than those in the Generate condition. Performance in the Context condition is generally intermediate between the other two conditions, except in free recall. Thus, these data suggest that the transfer appropriate processing approach provides the better interpretation of dissociation experiments than does the postulation of memory systems.

CONCLUDING COMMENTS

The argument put forward here is that dissociations among measures of retention can be best interpreted within a framework emphasizing the transfer appropriate procedures between study experiences and test occasions (Kolers & Roediger, 1984; Morris et al., 1977). Further, the distinction between data-driven and conceptually driven processing (Jacoby, 1983) is useful in describing differences among retention tests. Because these ideas are relatively new, some criticisms and qualifications are in order, as well as descriptions of future research that would be useful in testing these ideas.

First, how can one distinguish between data-driven and conceptually driven memory tests? What converging operations can be proposed? In these initial stages of inquiry a bootstrapping operation has been employed; that is, tasks that are sensitive to changes in surface features of information at study (modality, language, etc.) are said to be data-driven, whereas tasks that are relatively insensitive to such manipulations are classed as conceptually driven. A converging operation can be specified: Conceptually driven tasks should be greatly affected by study manipulations that vary elaborative processing, such as generating rather than reading items, or generating images to words rather than their rote repetition, etc. These same manipulations should have little effect, or even an opposite effect, on data-driven tests (depending on how the materials are presented at study). Thus, two converging operations can be proposed to distinguish data-driven from conceptually driven memory tests. The former should be greatly affected by changes in the surface form of studied information, but little affected by elaborate operations, and conversely for conceptually driven tests.
A second refinement, to expand on a remark made earlier, is that the data-driven/conceptually driven distinction should properly be considered as endpoints on a continuum rather than a strict dichotomy. Most memory tests will likely have components of each type of processing. For example, some theorists believe that recognition can be accomplished either by a fast-acting judgment process or a slower, more reflective deliberation usually called “memory search” (e.g., Atkinson & Juola, 1974; Mandler, 1980). In Jacoby’s (1983) terms, the former would be aligned with data-driven and the latter with conceptually driven processing. However, the balance between these two can be changed by test manipulations (Johnston, et al., 1985).

The same combination of influences is likely to be found on other tasks. The primary tasks that have been used to measure “retention without awareness” are (a) perceptual identification, (b) visual or auditory thresholds, (c) word fragment completion, (d) word stem completion, and (e) reading inverted text. These measures show priming effects in amnesics of roughly the same magnitude as found in normal subjects (except [b], which has not been examined to our knowledge). We would class all these as primarily data-driven tasks, but differences among them are found as a function of input modality. For example, auditory presentation of material produces no priming on measures of perceptual identification (Jacoby & Dallas, 1981, Experiment 6) or visual thresholds (Morton, 1979; Winnick & Daniel, 1970), but does produce facilitation in word stem completion (Graf, et al., 1985), word fragment completion (Roediger & Blaxton, 1987) and the reading of inverted text (Kolers, 1975).

Obviously, then, using cross-modal priming as an index of how data-driven a task is leads to the conclusion that perceptual identification and visual duration thresholds are more data-driven than are word stem completion, word fragment completion and reading inverted text. It would be of interest to see if these tasks order themselves the same way in detecting other changes in form of presentation, such as language of presentation for bilinguals, or verbal and pictorial representations. As yet, data are not available, even across experiments, to make meaningful comparisons on these dimensions. However, from our own experiments we can see indications that word fragment completion, when used as a memory test, is not completely data-driven. Besides cross-modal priming (Roediger & Blaxton, 1987), pictures produce reliable (if small) priming effects (Weldon & Roediger, 1987), as do instructions to form images of words’ referents (Blaxton, 1985). The large cross-modal priming effect seems to indicate the role of a lexical representation (on fragment completion) that is shared by items presented visually and auditorily. The effects from other modes of presentation may indicate small amounts of conceptual priming.

A further clarification is in order in describing tasks as data-driven. The meaning attached to this statement is that they are data-driven when used as tests of retention for recent episodes by presenting isolated items. We do not mean that no conceptual processes are ever used in perceptual identification, fragment
completion, or (certainly) reading inverted text. For example, if appropriate contexts are given, visual duration thresholds for words can be increased or decreased (Tulving & Gold, 1963). Surely similar conceptual manipulations would affect the other measures, whether these were used to measure retention for recently presented episodes or for "semantic memory" information. Unpublished data collected by the first author in collaboration with Bradford Challis and Valerie Ludwick show that this assumption holds true for word fragment completion, because modification of fragments by synonyms, associates and the like greatly enhances subjects' ability to complete the fragments, relative to the condition in which fragments are presented without context.

Another important facet of these tasks that measure retention without awareness is the instructions given to the subjects. Following the tradition of research with amnesics, these tasks are often not presented as retention tests. Rather, because subjects can potentially complete each task without recourse to memory for the episode, they are simply told before the test phase to identify the words or to complete the fragments, etc., as well as possible. In fact, different effects may be found when subjects are told to use the cues given at test to produce previously studied material, as Graf and Mandler (1984, Experiment 3) have shown in the case of word stem completion. Another obvious direction for future research is to see how sensitive these tasks are to manipulation of subjects' expectancies.

Other classifications have been proposed to account for differences in tasks that measure retention with and without awareness. These include the episodic/semantic distinction (Tulving, 1983), the procedural/declarative distinction (Cohen & Squire, 1980), and the description of the tasks as measuring implicit versus explicit memory (Graf & Schacter, 1985). These typologies partly overlap those of data-driven and conceptually driven processing as described by Jacoby (1983) and developed here. A primary task for future research is to discover which of these distinctions is most useful for understanding the dissociations among measures of retention. The ideas put forward here have the advantage of (a) emphasizing processing components of tasks rather than metaphysical systems lurking in the mind, (b) permitting operationalization of the terms, and (c) providing relatively straightforward means of falsification of the ideas if converging operations do not bear out the ordering of tasks in terms of their data-driven and conceptually driven components.

ACKNOWLEDGMENTS

We would like to thank Robert R. Hoffman, James H. Neely, Endel Tulving, and Mary Susan Weldon for their comments on a prior draft of the manuscript. The research reported here was supported by Grant RO1 HD15054 from the National Institute of Child Health and Human Development.