JOURNAL OF MEMORY AND LANGUAGE 29, 433–454 (1990)

Toward a Generate/Recognize Model of Performance on Direct and Indirect Tests of Memory

LARRY L. JACOBY AND ANN HOLLINGSHEAD

McMaster University

In two experiments, we compared performance on different tests of memory. Word stems were presented to be completed as an indirect test of memory. For a direct test of memory, the same stems were presented as cues for recall of earlier-presented words. Subjects in a generate/recognize test condition generated a completion for each word stem and then judged whether the generated word was presented earlier. Interactions between test conditions and a prior processing, and a materials variable were successfully predicted by a generate/recognize model of recall that postulates two bases for memory decisions. Cued-recall differs from stem-completion performance in that recognition processes are involved in the former but not the latter task. However, subjects do sometimes output words as recalled without doing a recognition-memory check. This failure to reliably do a recognition check results in recall of unrecognized words.

An important problem for theories of memory is to explain how memory for an event is influenced by prior knowledge. The most popular solution to this problem is derived from Bartlett's (1932) proposal of reconstructive memory processes. For example, a dominant theme in investigations of memory for prose has been to show that errors in recall can be explained as originating from a schema that was activated during study and used at test to reconstruct the original prose (e.g., Cofer, Chmielewski, & Brockway, 1976; Dooling & Christiaansen, 1977; Thorndyke, 1977). A difficulty for a schema approach is that although memory for prose sometimes appears to rely on reconstructive memory processes, at other times it appears to rely on reproductive memory processes in that very few recall errors are observed (e.g., Hasher & Griffin, 1978). It is tempting to describe this variation in recall accuracy in terms of a generate/recognize model of recall (e.g., Kintsch, 1970). Recall errors would then be said to occur when subjects fail to check the accuracy of their reconstructions. However, work done by Tulving and his colleagues to show the importance of encoding specificity (e.g., Tulving & Thomson, 1973) has discouraged the development of generate/recognize models of recall.

Comparisons of performance on indirect and direct tests of memory (e.g., Richardson-Klavehn & Bjork, 1988) potentially allow one to gain some insight into the relation between reconstructive and reproductive memory processes. When given an indirect test of memory, subjects are not asked directly to report on memory for a prior event, but, rather, are required to engage in a task that can reveal effects of prior experience. For example, subjects might be presented with a word stem (e.g., mot__) and told to complete the stem with the first word that comes to mind (e.g., motel). A finding that the earlier presentation of a word increases the probability of its being given as a completion constitutes indirect evidence of memory for prior presentation of the word (e.g., Graf & Mandler, 1984). Recognition and recall are direct tests of memory because the subjects are instructed

This research was supported by a grant to the first author from the National Science and Engineering Research Council. The authors thank F. I. M. Craik, H. L. Roediger, E. J. Shoben, and I. Begg for their comments on an earlier version of this article. Requests for reprints should be sent to Larry L. Jacoby, Department of Psychology, McMaster University, Hamilton, Ontario, Canada L8S 4K1.

0749-596X/90 $3.00

to report on an event in their personal history, such as the presentation of a word in a list. With reference to a generate/recognize model, performance on an indirect test of memory can be treated as reflecting influences of memory for an earlier event on reconstructive or generation processes and as differing from performance on a direct test in that the latter includes additional reproductive or recognition processes.

In two experiments, we compared effects of prior experience on stem-completion performance with effects on cued-recall performance given word stems as cues. The results of those experiments will be described in terms of a generate/recognize model that differs in important ways from the class of generate/recognize models of recall that was successfully rejected by Tulving and his colleagues (e.g., Tulving & Thomson, 1973; Watkins & Gardiner, 1979). Before introducing our experiments, we briefly describe the generate/recognize model that guided our research.

Generation Processes: An Episodic View of Knowledge

An important difference between the generate/recognize model that we propose and earlier models concerns the form of knowledge said to underlie effects of prior experience on generation processes. According to earlier models (Anderson & Bower, 1972; Bahrick, 1970; Kintsch, 1970), the generation of candidates for recall relies on knowledge represented in the form of an abstract, fixed associative network that is used in an invariant fashion across situations. Similarly, effects of prior experience on performance of an indirect test of memory have been explained as produced by the activation of an abstract representation such as a logogen or a schema (e.g., Graf & Mandler, 1984). In contrast, we (e.g., Jacoby, 1983; Jacoby & Brooks, 1984) hold that memory for prior episodes is responsible for effects of prior experience on generation processes, as measured by performance on an indirect test of memory, even when people are unable to recognize a generated item as earlier presented. Our view shares assumptions with "exemplar" or "instances" accounts of concept learning advanced by Brooks (1978, 1987), Medin (Medin & Schaffer, 1978), and Hintzman (1986). By those accounts, variability in performance across situations results from the differential contribution of memories of particular instances of a concept or type of event and is greater than could be produced by reliance on some general, abstract representation. It is variability across situations in memory performance as measured by an indirect test that we use to argue that generation processes reflect memory for prior episodes.

The claim that generation processes are influenced by memory for prior episodes partially disarms Tulving's criticism of generate/recognize models of recall (Tulving, 1976). This is because the form of his attack on those models was to show effects of encoding specificity on cued-recall performance. For example, Thomson and Tulving (1970) showed that when an item was studied in the context of a weak associate (e.g., ground COLD), that weak associate was a more effective cue for later recall than was a strong associate of the target item. That result is damaging to earlier generate/recognize models, because one would expect strong associates to invariably do more to aid recall than would weak associates if cued recall was accomplished by using an abstract, fixed associative network to generate candidates for recall. However, the result can be explained by a generate/recognize model if it is assumed that presentation of an item in the context of a weak associate influences generation as well as recognition processes underlying later cued-recall performance. But, then, to justify a claim of effects on both types of processes, one must demonstrate effects of prior presentation of an item on generation processes that are separate from effects on recognition processes. Also, if one is to hold an episodic view, it must be shown that effects on generation processes are specific to memory for prior episodes rather than being more general effects of the sort that would arise from the activation of an abstract representation.

Akin to encoding-specificity effects, performance on indirect tests of memory reflects the match between the study and the test processing of an item (Allen & Jacoby, in press; Jacoby, 1983; Roediger & Blaxton, 1987). For example, Allen and Jacoby (in press) presented words in their normal form to be read or as anagrams to be solved (e.g., dowry vs yodrw) and then assessed memory by means of a perceptual-identification or a recognition-memory test. (To make the anagrams easier to solve, two letters were underlined to indicate that their positions in the anagram were the same as in the solution word.) For the memory tests, words were presented in their normal form. Results showed that words read earlier, as compared with those presented as anagrams, were *more* likely to be perceptually identified but were *less* likely to be recognized as earlier presented. Both effects were large and highly reliable. The finding of an opposite effect of a manipulation of prior processing on an indirect as compared with a direct test of memory is important in showing an effect of memory for prior episodes on generation processes that is separate from an effect on recognition processes.

The finding that the manipulation of prior processing had an opposite effect for the two types of tests can be explained as reflecting the importance of the compatibility of study and test processing (e.g., Jacoby, 1983; Roediger & Blaxton, 1987). For the perceptual-identification test, it is the match between prior presentation and test in the visual details of an item along with the match in perceptual processing that are important for performance, producing an advantage for words that were read over those presented as anagrams. In contrast, an activation account would have to predict that the two study conditions would produce equivalent stem-completion performance. This is because an activation view holds that effects of prior experience are mediated by an abstract representation and, consequently, should be general, and not restricted by the details of memory for a prior presentation of an item. The advantage in recognition-memory performance of anagram over read words is similar to the "generation" effect (e.g., Jacoby, 1978; Slamecka & Graf, 1978) and, presumably, shows that recognition memory benefits from more extensive or elaborative processing. To anticipate, in the experiments that are to be reported, we used the finding that the read vs. anagram manipulation has opposite effects on generation as compared with recognition processes to gain support for a generate/recognize model that holds that the reliance of cued-recall performance on recognition processes varies across situations.

Given that generation processes are influenced by memory for prior episodes, a generate/recognize model might not be needed to explain differences between stem-completion and cued-recall performance. Any effect of test instructions might be totally produced by a difference in generation processes, meaning that recognition processes are not involved in cued-recall performance. However, we will show that when word stems are provided as cues, a manipulation of cued-recall vs. stem-completion instructions has little, if any, effect on generation processes. Rather, the effect of giving cued-recall instructions is to add recognition-memory processes to the generation processes required for stem-completion performance.

The encoding-specificity principle is not antagonistic to a generate/recognize model of recall. If transfer effects are quite specific, as described by the encoding-specificity principle, then an item coming promptly to mind in response to a question about memory makes it likely that the item is one that was previously presented. However, subjects need not be aware of the match between encoding and retrieval for that match to affect text performance. The compatibility of study and test processing is important for indirect as well as for direct tests of memory (e.g., Allen & Jacoby, in press). Also, an item can come promptly to mind for reasons other than its earlier presentation. For example, a particular response may come to mind because the task is so tightly constrained that no other response is possible, as would be true if a word stem with only one possible completion was presented as a cue for recall. Consequently, a recognition-memory check in the form of an attempt to recollect the prior presentation of an item is sometimes a necessary addition to the generation phase to avoid falsely recalling items as previously presented.

Recognition Processes: Multiple Bases for Judgments

The evidence taken by Tulving and his colleagues as most damaging to generate/recognize models of recall was the finding of cued recall of unrecognizable words (e.g., Tulving & Thomson, 1973). They argued that, contrary to their results, generate/recognize theories must predict that unrecognized words will not be recalled. This is because generate/recognize models (e.g., Anderson & Bower, 1972; Bahrick, 1970; Kintsch, 1970) treated recognition failure as implying the absence of an adequate representation in memory. However, one way that generate/recognize models can predict recognition failure for recallable words is by assuming that the decision criterion differs for a recognition-memory test as compared with a recall test (Kintsch, 1978). Kintsch quoted Müller as support for the claim of a difference in criteria. By Kintsch's translation (1978, p. 472), Müller (1913) stated: "the manner of retrieval itself may play a role in the editing process: If an item is retrieved promptly, it may be admissible even though its familiarity value is low." This means that, contrary to the usual assumptions, the recognition criterion for recall might be lower than that for a test of recognition memory.

Müller's description of decision criteria for recall can be slightly modified to describe a generate/recognize model for recall that is similar to dual-process models of memory performance (e.g., Atkinson & Juola, 1974; Glucksberg & McCloskey, 1981; Jacoby & Dallas, 1981; Mandler, 1980; Reder, 1987). Indeed, a claim made by Jacoby and colleagues that the feeling of familiarity reflects the operation of a fluency heuristic (e.g., Jacoby & Dallas, 1981; Jacoby, Kelley, & Dywan, 1989a; Kelley, Jacoby, & Hollingshead, 1989) is closely in line with Müller's description of criteria used for recall. The notion is that neither recognition nor recall tests are "pure" as regards the basis for memory decisions. For both types of tests, assessing the relative fluency of identification or of generation serves as an alternative to recollection as a basis for judging an item as old. Our claim of two bases for memory judgments is a claim of a qualitative difference between criteria, whereas Kintsch's (1978) claim of a difference in criteria referred to a quantitative difference in criterial memory "strength." By our dual-process model, cued recall of unrecognizable items arises from subjects' failure to carry out a recognition check when the generation of a candidate for recall is relatively fluent.

Interactions with manipulations of prior processing can be used to index the involvement of recognition processes in cued-recall performance. In this vein, it is particularly useful that presenting words to be read vs. as anagrams to be solved has an effect on recognition-memory performance that is opposite to the effect on generation processes (Allen & Jacoby, in press). Suppose that, although given cued-recall instructions, subjects completed word stems with the first word that came to mind. One would than expect higher performance for words that were earlier read as compared with words earlier presented as anagrams, because of the advantage in generation processes held by read words. The failure to rely on recognition-memory processes would result in the cued recall of words that could not be recognized as old on a test of recognition memory, particularly when words were read when earlier presented. If the situation were changed in a way that put heavier weight on recognition-memory processes, the effect would be to reduce or, perhaps, reverse the difference in cued-recall performance for anagram vs. read words. Under those circumstances, the probability of cued recall for words earlier presented as anagrams would likely exceed that of words that were earlier read because anagrams hold an advantage in recognition-memory performance. The finding of opposite effects on cued recall as a function of differential reliance on recognition processes depends on prior-presentation effects on recognition-memory processes being large enough to reverse effects on generation processes.

The probability of an earlier presented word being given as a response is generally higher when a word stem or fragment is presented as a cue for recall than it is when subjects are instructed to complete the stem or fragment with the first word that comes to mind (Graf & Mandler, 1984; Graf, Squire, & Mandler, 1984; Nelson, Canas, Bajo, & Keelean, 1987; Nelson, Keelean, & Negrao, 1989; Weldon, Roediger, & Challis, 1989). That effect might result from an influence of test instructions on generation processes that is independent of any differential involvement of recognition-memory processes. However, an effect of that sort is not necessary to produce superiority of cued-recall instructions when stems or fragments allow multiple completions, as was true in most earlier experiments. When multiple completions are possible, recognition processes brought into play by cued-recall instructions can serve to select an earlier presented word from among alternative completions. This selection makes it more probable that the old word will be given as a response with cued-recall than with stem-completion instructions.

In order to demonstrate an effect of test instructions on generation processes that is independent of any effect on recognition processes, it would be necessary to show that cued-recall performance surpasses stem-completion performance when stems allow only a single completion. Under those circumstances, an involvement of recognition processes in cued recall would have no effect if recognition performance were perfect and, otherwise, would make it less likely that an old word would be given as a completion with cued-recall than with stem-completion instructions. Weldon et al. (1989) compared cued-recall with fragment-completion performance using fragments that allowed only a single completion. The results of their experiments failed to show an advantage of cued-recall over fragment-completion performance and, so, failed to provide evidence of effects of test instructions on generation processes that are independent of effects on recognition-memory processes. Their data are consistent with the possibility that cued-recall differs from fragment-completion performance only in that the former involves recognition-memory processes that are not involved in the latter, a generate/recognize model of recall.

Other data from the experiments by Weldon et al. (1989) can also be interpreted as providing support for a generate/recognize model of recall. The test lists in their experiments included fragments that could not be completed with any of the earlier presented words. When cued-recall instructions were given, subjects produced false recalls by completing those baseline fragments, "recalling" words that were not earlier presented. False recalls would be expected if subjects failed to do a recognition-memory check before outputting words that came promptly to mind as a completion for a fragment. In our experiments, the test lists that we used included word stems that could only be completed with a new word so as to allow us to examine effects on the probability of false recall.

The Experiments

Our first experiment compared cued-recall with stem-completion performance and with performance in a generate/recognize test condition. In the latter condition, subjects completed each word stem to produce a five-letter word, saying the first suitable word that came to mind, and then judged whether the word they had just given as a completion was presented earlier. The probability of generating and recognizing old words was compared with performance in the cued-recall test condition. The test list included stems that could not be completed with any of the earlier presented words. Those stems provided a baseline measure for the stem-completion condition and served to measure the probability of false recall in the other two test conditions. Word stems that were used comprised the first three letters of a five-letter word and allowed either multiple completions or could be completed with only a single five-letter word. We constrained subjects to completing stems with five-letter words because we thought that made the completion task easier than it would be otherwise, and, consequently, might produce a higher probability of false recall. Word stems along with two blanks to indicate missing letters were also provided as cues for recall.

A generate/recognize model of recall that postulates two bases for memory decisions predicts that performance in the cued-recall condition will reflect a mix of performance in the other two test conditions. This is because it is assumed that items that are fluently generated are not subjected to a recognition check before being output as recalled. Performance on those items should be functionally equivalent to performance in the stem-completion condition, whereas performance on items that are subjected to a recognition check before being output should be functionally equivalent to performance in the generate/recognize test condition. When making predictions, we assume that the test conditions are identical with regard to generation processes and differ only to the extent that they involve recognition-memory processes. Results for the single-completion stems were of major interest, because they have the potential to reveal violations of that assumption. As argued earlier, when stems allow only a single completion, the probability of a word being given as a response for a cued-recall test cannot be higher than that for a stem-completion test unless test instructions influence generation processes.

Words that served as old words in the second phase of the experiment were earlier presented either in their normal form to be read or in anagram form to be solved. Solving an anagram to produce a word as compared with reading a word produces opposite effects on generation as compared with recognition-memory processes (Allen & Jacoby, in press). For stem-completion performance, the read words were predicted to hold an advantage over the solved words, because the earlier reading of a word does more to aid later generation processes. Because of added recognition-memory processes, it was predicted that the advantage of read over anagram words would be eliminated or, perhaps, even slightly reversed when cued-recall rather than stem-completion instructions were given, and that the advantage of anagram over read words would be further increased in the generate/recognize test condition. That interaction between prior presentation and test conditions was predicted because of differential reliance (across test conditions) on recognition-memory processes, and because words presented as anagrams hold an advantage over read words in recognition-memory performance. It was thought unnecessary to add a fourth test condition to examine that recognition-memory advantage. The advantage was earlier shown by Allen and Jacoby (in press). Also, performance in the generate/recognize test condition can be examined to determine whether generated words that were earlier presented as anagrams were more likely to be recognized as old than were generated words that were earlier read.

The probability of completing word stems that can only be completed with a new word should decline as one goes from the stem-completion to the cued-recall to the generate/recognize test condition, because increased reliance on recognition-memory processes should make it more likely that subjects will avoid falsely recalling new words. However, it is unnecessary to use the probability of false recall to correct for differences in guessing when comparing performance on word stems that can be completed with an old word (cf. Weldon et al., 1989). The interaction between read vs. anagram and test conditions is predicted for uncorrected probabilities, and is expected to arise because of differences among test conditions in their reliance on recognition-memory processes.

Generate/recognize theories of recall were meant originally to explain interactions between characteristics of studied material and type of test. For example, high-frequency words as compared with low-frequency words hold an advantage in recall but are less likely to be recognized. This pattern of results can be interpreted as showing that high-frequency words hold an advantage for generation processes but are at a disadvantage, as compared with low-frequency words, for recognition-memory processes (Gregg, 1976). We manipulated frequency of words in the language in our experiment and predicted a significant interaction between frequency in the language and test conditions for reasons that are similar to those used to predict an interaction between read vs. anagram and test conditions. Because of the effect of frequency on generation processes, the probabilities of stem completion and of false recall were predicted to be higher for medium- as compared with low-frequency words. Across test conditions, increases in reliance on recognition-memory processes were expected to act to the advantage of low- over medium-frequency old words, because low-frequency words are more likely to be recognized as earlier presented. As argued for the interaction of read vs. anagram with test conditions, the interaction between frequency in the language and test conditions can be used to index the involvement of recognition processes in cued-recall performance.

A second experiment was very similar to the first except that we gave subjects in the cued-recall test condition feedback so as to discourage false recall. The notion was that encouraging subjects in the cued-recall condition to rely more heavily on recognition-memory processes would increase the similarity between that condition and the generate/recognize test condition.

Experiment 1

Method

Subjects. The subjects were 72 volunteers from an introductory psychology course at McMaster University who served in the experiment for course credit. Twenty-four subjects were randomly assigned to each of three between-subject test conditions (Stem Completion, Generate/Recognize, and Cued Recall). Subjects were tested individually.

Design and materials. We selected a pool of 100 five-letter nouns as stimuli. The first three letters (word stems) of each of those words were unique within the set of words, but not always unique in the language. The stems of 50 of the words could be completed with four or more different five-letter words (e.g., for--, foray, force, forge, forgo, forte, forth, forty, forum), whereas the other 50 stems could be completed with only one five-letter word in English (e.g., oas--, oasis). An additional criterion for the selection of words was the ability to construct an anagram with only one solution from them, given the restriction that the position of the second and fourth letters of the anagram was the same as in the selected word (e.g., *yodrw*, dowry). Selected words were further divided with regard to frequency in the language; 50 medium-frequency (10 to 49 per million) and 50 low-frequency (1 to 5 per million) words as scaled by Thorndike-Lorge (1944). Each of the four possible combinations of frequency in the language and number of possible completions for word stems was represented by 25 words. Materials are presented in Appendix A.

For each set of 25 words, five were designated as new words. Those words were never presented during the first phase of the experiment but served as a source of stems presented at the time of test. The new items served as a baseline measure for the stem-completion condition and as a measure of false recall for the other two test conditions. The remaining 20 words in each set were divided evenly between those presented as anagrams and those presented in their normal form to be read in the first phase of the experiment. This resulted in an 80-word study list that included 40 words to be read and 40 anagrams to be solved. Two formats were constructed such that words presented to be read in one format were presented as anagrams to be solved in the other format. New words were not counterbalanced with old words. Doing so was considered unnecessary because primary interest was in the interaction of prior presentation (read vs. anagram) with test conditions. A 100 word-stem test list included 80 stems that could be completed by words presented earlier (40 read, 40 anagrams) and 20 stems that could not be completed by any of the earlier presented words. The presentation order of items for both study and test lists was random with the restriction that not more than three items representing the same combination of conditions could be presented in a row.

Procedure. An Apple IIe computer interfaced with a monochrome green monitor was employed to present the stimuli. The character size of the stimuli was approximately 5.7 × 6.6 mm. Words and word stems were presented in lower case letters in the center of the screen.

In the first phase of the experiment, subjects were required to solve anagrams and to read words. They were informed that words would sometimes be presented in their normal form and that their task was to read those words out loud as quickly as possible. Subjects were told that other words would be presented as anagrams with the second and fourth letters underlined, and that those underlined letters were in correct positions with reference to the solution word. It was explained that the underlining of letters was done to make the anagrams easier to solve and to allow only one solution for each anagram. If the word said aloud by the subjects was the correct solution for a presented anagram, the experimenter pressed a key to initiate presentation of the next item. Otherwise, the subjects were informed of their error and were required to continue attempting to solve the anagram. A maximum of 30 s was allowed for the solving of each anagram. Once that time limit elapsed, signaled by a beep produced by the computer, the experimenter told the subjects the solution word. Subjects were encouraged to compare the solution word with the anagram to be certain that the solution given was the correct one. After each item had been presented and either read or solved, the experimenter pressed a key to initiate the next trial. Subjects were led to believe that times were being recorded both for reading the words and for solving the anagrams, and that the reading times were to be used as a baseline for interpreting the times to solve anagrams. In fact, times were not recorded. No mention was made of a retention test.

In the test phase, word stems were presented as the initial three letters of a word followed by two dashes. Instructions differed for the between-subject test conditions (Stem Completion, Generate/Recognize, and Cued Recall). In the stem-completion condition, subjects were instructed to complete the word stem by filling in the dashes to form the first five-letter word that came to mind. No proper names or plurals were allowed. If the subjects' response met these criteria, the experimenter pressed a key to remove the word stem from the screen and then pressed another key to initiate presentation of the next word stem. Otherwise, subjects were informed of their error and were told to attempt to give a satisfactory completion for the word stem. Subjects were allowed a maximum of 20 s to complete each stem. If after 20 s the word stem had not been completed, a beep sounded at which time the experimenter pressed a key to clear the screen and then another key to initiate the next trial.

In the generate/recognize condition, subjects were required to make a recognition-memory decision for the word they produced as a completion for a stem. The procedure for the stem-completion portion of the task was the same as described for the stem-completion condition. After a completion was given, the experimenter pressed a key and the message "old/new?" appeared on the screen. That message prompted subjects to make a recognition-memory decision for the word that they had just generated as a completion for a stem. If they remembered encountering the word in the earlier part of the experiment, they were to respond old. Otherwise, they were to respond new. Subjects were encouraged to treat the completion and recognition-memory tasks as being totally separate. They were instructed to complete each word stem with the first word that came to mind, without considering whether the word had been presented earlier. Only after a word had been given as a completion, were they to make their recognition decision.

The subjects in the cued-recall condition were instructed to use the word stems as cues for recall of words read or solved in the first part of the experiment. As soon as they recalled a word given a stem as a cue, they were to report the word out loud. Once they had given their response, the experimenter pressed a key to clear the word stem from the screen and then pressed another key to initiate presentation of the next word stem. Subjects were informed that there was a 20-s time limit for responding, and that a beep would sound when time was up. It was further explained that they were not required to give a response for every stem. They could simply say "pass" any time during the 20-s interval if they felt they could not complete the stem with an earlier-presented word. It was stressed that they were to complete the stems only with words they could recall as presented in the earlier part of the experiment.

The significance level for all tests was set at $p < .05$.

Results and Discussion

The probability of solving anagrams in the first part of the experiment was .88. For the test phase, because of the complexity of the design and because of differences in predictions that can be made when word stems allow only single rather than multiple completions, results were analyzed separately for the two types of word stem. For each type of word stem, we first report analyses done on results from the generate/recognize test condition to examine effects on recognition-memory performance and to reveal any effect of test instructions on generation processes. Next, we compare cued-recall performance with performance in the other two test conditions.

Single Completions

Effects on recognition memory and on generation processes. Recognition-memory performance in the generate/recognize condition was analyzed by examining effects on the probability of correctly calling an item old, given that it had been generated. The results of that analysis showed a recognition advantage for words presented earlier as anagrams over those earlier read (.82 vs. .56), $F(1,23) = 121.53$, MSe =

.013. Also, low-frequency words were more likely to be correctly recognized than were medium-frequency words (.74 vs. .64), $F(1,23) = 7.72$, $MSe = .030$. A significant interaction between frequency in the language and prior presentation, $F(1,23) = 4.68$, $MSe = .024$, resulted from the recognition advantage of low- over medium-frequency words being larger when words were earlier read (.65 vs. .48) than it was when words were earlier solved (.84 vs. .81). The effect of frequency in the language and the read vs. anagram effect are consistent with other results in the literature. The interaction between the two factors presumably reflects the fact that recognition memory was near ceiling when words were earlier presented as anagrams, restricting the size of the effect of frequency in the language.

The task of recognition memory in the generate/recognize condition might have produced differences in generation processes in that condition compared with the stem-completion condition. To check that possibility, performance in the generate/recognize condition was analyzed further to examine effects on generation processes separately from effects on recognition processes. For the generate/recognize condition, the probabilities of completing a stem with words earlier presented as anagrams, words earlier read, and new words were: .91, .91, and .72, respectively. An analysis comparing stem-completion performance in the generate/recognize condition with performance in the stem-completion condition revealed neither a significant main effect nor any significant interaction involving test condition. That is, requiring a recognition test did not significantly influence generation processes.

Comparisons with cued-recall performance. Results obtained using single-completion stems, collapsed across frequency in the language, are displayed in Table 1. For the separate test conditions, the probabilities refer to the probability of stem completion, the probability of cued recall, and the probability of completing a stem and claiming to recognize as old the word given as a completion. For stems that could only be completed with a new word, false recall was shown by giving the new word as a response for cued recall, or by generating and falsely recognizing the new word (see the third column of Table 1).

TABLE 1

PROBABILITIES OF STEM COMPLETION, CUED RECALL, AND STEM COMPLETION PLUS RECOGNITION FOR *SINGLE*-COMPLETION WORD STEMS

	Study condition		
Test condition	Anagram	Read	New
Experiment 1			
Stem completion	.89	.91	.72
Cued recall	.82	.77	.46
Generate/recognize	.75	.51	.13
Experiment 2			
Cued recall	.79	.56	.11
Generate/recognize	.72	.45	.09

The probability of a new word being falsely recalled in the cued-recall condition was quite high, but substantially lower than the probability of a new word being generated as a completion in the stem-completion condition (.46 vs. .72), $F(1,46) = 23.14$, $MSe = .069$. That result is understandable if some, but not all, generated words were subjected to a recognition check before being output for cued recall. The probability of false recall was much higher than the probability of generating and falsely recognizing a new word as indexed by performance in the generate/recognize condition (.46 vs. .13), $F(1,46) = 55.44$, $MSe = .047$. Similarly, when stems could be completed with an earlier presented word, the probability of cued recall for old words (.80) was lower than the probability of stem completion (.90), $F(1,46) = 17.03$, $MSe = .029$, but higher than the probability of generation and recognition of an old word (.63), $F(1,46) = 22.12$, $MSe = .057$. Effects of test condition for both new and old words can be explained as showing that the involvement of recognition-memory processes increases as one goes from stem completion, to cued recall, to the generate/recognize test condition.

Interactions between form of prior presentation (anagram vs. read) and test conditions also point toward the differential involvement of recognition processes. Across test conditions, an increase in the involvement of recognition processes was reflected by an increase in the advantage of anagram over read words. Words earlier presented as anagrams held a slight advantage over words earlier read in cued-recall (.82 vs. .77) but not in stem-completion performance (.89 vs. .91), $F(1,46) = 3.25$, $MSe = .016$, $p < .10$ for the interaction. The advantage of anagram over read words in the generate/recognize condition (.75 vs. .51) was larger than that in the cued-recall condition (.82 vs. .77), $F(1,46) = 14.27$, $MSe = .027$.

Medium- as compared with low-frequency words were expected to be more likely to be generated but less likely to be recognized as previously presented. Results (see Table 2) were generally in accord with our predictions. The effect of frequency in the language on generation processes was revealed by performance on stems that could be completed only with a new word. Comparing cued-recall and stem-completion performance, subjects in both test conditions were more likely to respond with a medium- than with a low-frequency (.70 vs. .48) new word, $F(1,46) = 28.37$, $MSe = .043$. Similarly, a comparison of performance in the cued-recall and the generate/recognize conditions showed that, across conditions, medium-frequency words were more likely to be falsely recalled (.40) than were new low-frequency words (.19), $F(1,46) = 30.81$, $MSe = .032$. Those effects can be attributed to an influence of frequency in the language on generation processes. When stems could be completed with an old word, a comparison of performance in the cued-recall and the generate/recognize conditions revealed a significant interaction between test condition and frequency in the language, $F(1,46) = 7.44$, $MSe = .030$. Examining that interaction, low-frequency words were more likely to be generated and correctly recognized than were medium-frequency words (.66 vs. .59), whereas medium-frequency words were more likely to be recalled than were low-frequency words (.82 vs. .76). The form of that interaction is consistent with the predicted greater involvement of recognition processes in the generate/recognize, as compared with the cued-recall, test condition.

TABLE 2
EFFECTS OF FREQUENCY IN THE LANGUAGE ON THE PROBABILITIES OF STEM COMPLETION, CUED RECALL, AND STEM COMPLETION PLUS RECOGNITION

	Word type			
	Old frequency		New frequency	
	Low	Medium	Low	Medium
Stem completion	.89	.91	.62	.82
Cued recall	.76	.82	.33	.58
Generate/ recognize	.66	.59	.05	.21

Multiple Completions

Effects on recognition memory and on generation processes. Recognition-memory performance was higher for words earlier presented as anagrams than it was for words earlier read (.79 vs. .46), $F(1,23) = 49.78$, $MSe = .053$. Unlike results for single-completion stems, the effect of frequency in the language on recognition-memory performance was not significant. That lack of an effect is probably because subjects often completed stems with words other than those that we chose as target words when stems allowed multiple completions, but could not do so when stems allowed only a single completion. We could examine effects on recognition-memory performance only when stems were completed with an old word, meaning that analyses were based on a larger number of observations for single-completion as compared with multiple-completion stems.

The probabilities of stem completion in the generate/recognize condition were .36, .40, and .94 (.22) for anagram, read, and new words, respectively. The number in parenthesis for the new words is the probability for completions with words designated as target words. Although the interaction was not significant, the disadvantage in completion performance of anagrams compared with read words was somewhat larger in the stem-completion (.32 vs. .43) than in the generate/recognize condition, suggesting that the task of recognition memory might have had some small effect on generation processes.

Comparisons with cued recall. Probabilities of stem completion, cued recall, and completion and recognition are shown in Table 3. The numbers in parentheses in that table for new words are the probabilities of completion with a word designated as a target, using the same criteria to specify targets as for old words.

The manipulation of test instructions had an effect when stems could be completed only with a new word. For those stems, the probability of stem completion was much higher than was the probability of false recall (.94 vs. .48), $F(1,46) = 51.49$, $MSe =$ .049. Also, the probability of false recall was higher than the probability of generating and mistakenly recognizing a word as old (.48 vs. .14), $F(1,46) = 25.38$, $MSe =$ .108. Those effects parallel effects reported for stems that allowed only a single completion.

TABLE 3
PROBABILITIES OF STEM COMPLETION, CUED RECALL, AND STEM COMPLETION PLUS RECOGNITION FOR *MULTIPLE*-COMPLETION WORD STEMS

	Study condition		
Test condition	Anagram	Read	New
Experiment 1			
Stem completion	.32	.43	.94 (.17)
Cued recall	.36	.33	.48 (.07)
Generate/recognize	.28	.19	.14 (.04)
Experiment 2			
Cued recall	.34	.25	.09 (.04)
Generate/recognize	.28	.18	.10 (.03)

Note. Numbers in parentheses are the probabilities of completion with or false recall of words designated as target words.

An analysis comparing cued-recall with stem-completion performance revealed a significant interaction between read vs. anagram and test condition, $F(1,46) = 8.60$, $MSe = .025$. Words earlier read, as compared with words earlier presented as anagrams, were more likely to be given as a completion for a stem but were slightly less likely to be recalled (see Table 3). The advantage in stem-completion performance of read words over words presented as anagrams (.43 vs. .32) shows the importance for generation processes of the compatibility of the study and the test processing of an item (Allen & Jacoby, in press; Jacoby, 1983; Roediger & Blaxton, 1987) and, consequently provides evidence against an activation view of the effect of prior experience on stem-completion performance. An activation view would predict equivalent performance in the two test conditions, because activation of an abstract representation would produce general effects, not effects that are specific to the prior processing of an item.

The probability of responding with an old word when given a cued-recall test was higher than the probability of generating and recognizing an old word (.34 vs. .24), $F(1,46) = 22.57$, $MSe = .024$. For both types of tests, words presented earlier as anagrams were more likely to be given as a response than were words that were earlier read (.32 vs. .26), $F(1,46) = 9.19$, $MSe =$ .020. Although the interaction was not significant, $F(1,46) = 2.22$, the advantage for anagram over read words was somewhat larger in the generate/recognize condition than in the cued-recall condition (see Table 3). As was true for single-completion stems, results for multiple-completion stems can be explained as showing that reliance on recognition-memory processes differed among the test conditions.

From the perspective of test conditions, the probability of giving an old word as a response was higher for the stem-completion than for the cued-recall test when target words were earlier read (.43 vs. .33) but the opposite was true when target words were earlier presented as anagrams (.32 vs. .36). Graf and Mandler (1984) reported a similar interaction between level-of-processing and cued-recall vs. stem-completion performance. Interactions of that sort are easily understood if recognition processes are involved in cued-recall but not in stem-completion performance. Words that are unlikely to be recognized because they were only read or only processed to a shallow level are less likely to be output for a cued-recall than for a stem-completion test.

Summarizing, the results were generally consistent with predictions made by a generate/recognize model of recall that postulates two bases for memory decisions. This was true regardless of whether word stems allowed single of multiple completions. However, the results were somewhat more dramatic for the single-completion stems. The predicted effects of frequency in the language were observed for single- but not for multiple-completion word stems. As discussed earlier, that difference is probably because single- as compared with multiple-completion stems resulted in target words more often being given as a completion and, consequently, allowed a more sensitive test of differences. Analyses were necessarily restricted to performance when target words were given as a response. Also, when multiple-completion stems were used, there was a greater potential for item-selection effects that might offset other effects. For example, low-frequency words that were given as a completion may have been higher in frequency or differed in some other systematic way from low-frequency words that were not given as a completion, and differences of that sort may have diminished effects of frequency in the language.

Similar to results reported by Weldon et al. (1989), cued-recall did not surpass stem-completion performance when stems allowed only a single completion. In our experiment, there was little chance for this to happen because of near ceiling performance in the stem-completion condition. However, that lack of a difference does provide some evidence that the manipulation of test instructions did not have its effects by influencing generation processes separately from recognition-memory processes. Further evidence that test instructions had little, if any, effect on generation processes was provided by the lack of a significant interaction between prior presentation and completion performance in the stem-completion as compared with the generate/recognize test condition. Of course, both of those tests are rather weak ones, making it impossible for us to fully rule-out the possibility that test instructions influenced generation processes.

More important was our finding of significant interactions between test conditions and factors that have differential effects on generation and recognition processes. The form of those interactions provides strong evidence that the test conditions differed in their reliance on recognition-memory processes. Performance in the cued-recall test condition can be adequately described as a mix of performance in the other two test conditions. Subjects in the cued-recall test condition often output items as recalled without benefit of a recognition check. That this was so is shown both by the high probability of false recall in that test condition and by the interaction of read vs. anagram with cued-recall vs. generate/recognize test. It was, arguably, words that were fluently generated as a completion that were recalled without being checked for recognition. We do not have latency data to provide as direct evidence for the importance of fluency. However, the pattern of results is in line with the possibility that an assessment of relative fluency served as a basis for deciding whether to perform a recognition check. For example, medium-frequency words were more likely to be given as a completion and were also more likely to be falsely recalled than were low-frequency words.

Across test conditions, we observed recognition failure of recallable words (cf. Tulving & Thomson, 1973). The probability of recognizing a word that had been earlier read was .56, whereas the probability of cued recall for those words was .77. However, that pattern of results is consistent with our generate/recognize model of recall. The superiority of cued-recall over recognition-memory performance resulted from subjects' failure to reliably use recognition as a criterion for outputting words as recalled. As evidence that this was true, the probability of false recall was much higher than was the probability of generating and falsely recognizing a new word as old.

The probability of false recall was surprisingly high in our experiment as compared to the results of Weldon et al. (1989). That high level may have resulted because our word stems were easier to complete than were their word fragments. Alternatively, the restriction that stems must be completed with five-letter words may have reduced reliance on recognition processes for a reason other than any effect on ease of completing stems. It may be that imposing additional restrictions on items that are to be produced for cued recall makes it likely that subjects will check to determine whether those restrictions have been met at the expense of checking for recognition memory. Either way, this means that providing additional cues (restrictions) for recall can have the seemingly paradoxical effect of increasing the probability of a false alarm, an effect similar to that often taken as evidence for reconstructive memory processes.

Experiment 2

The probability of false recall was very high for subjects in the cued-recall condition in Experiment 1. For Experiment 2, details of the cued-recall test were changed in ways that were meant to reduce the probability of false recall by making it less likely that subjects would accept a generated word as old without completing a recognition check. Subjects in the cued-recall condition in Experiment 1 were not informed that the test list included word stems that could not be completed with an earlier presented word. In contrast, subjects in Experiment 2 were warned that some stems could not be completed with an old word and were told that those stems were included in the test list as "catch" trials to make sure that they were conforming to the cued-recall instructions. Also, when subjects mistakenly reported a new word as recalled in Experiment 2, a buzzer sounded to inform them of their error. Performance in the cued-recall condition was compared with that in a generate/recognize condition in Experiment 2. A stem-completion condition was not included in the experiment. The only other difference between the experiments was that, because of a minor change in the materials, the effects of frequency in the language were not analyzed in Experiment 2.

The changes in the procedures for the cued-recall condition were expected to bring the probability of false recall down to the level of the probability of generating and falsely recognizing a new word. This is because those changes were expected to increase the likelihood of subjects in the cued-recall condition using a recognition-memory check to qualify generated words for output as recalled. For the same reason, when stems could be completed with an old word, performance in the cued-recall condition was expected to be more similar to that in the generate/recognize condition in Experiment 2 than it was in Experiment 1. In particular, the heavier reliance on recognition processes for cued recall in Experiment 2 was expected to remove the significant interaction, observed in Experiment

1, between prior presentation (read vs. anagram) and cued-recall vs. generate/recognize test conditions.

Method

Subjects. The subjects were 48 introductory psychology students who served in the experiment for course credit. Twenty-four subjects were randomly assigned to each of two between-subject test conditions (Generate/Recognize and Cued Recall).

Materials and procedure. As a consequence of an item analysis performed on the stimuli used in Experiment 1, a total of 20 of those words were replaced: 10 words used as a source of single-completion word stems and 10 words used as a source of multiple-completion words stems. The reason for replacing words used as a source of single-completion stems was that those stems were very likely to be completed regardless of prior presentation (probability of .94 and up). For multiple-completion stems, words were replaced if their stems generated fewer than 4 or more than 10 different completion words. The replacement of words in the pool resulted in an unequal number of low- and medium-frequency words. Consequently, frequency in the language was not analyzed as a factor in Experiment 2 but, rather, was controlled by equating the words representing different conditions with regard to frequency in the language and by using multiple formats to rotate words through conditions. As in Experiment 1, new words were not counterbalanced with words presented to be read or with words presented in the form of anagrams.

There were two between-subject test conditions: cued recall and generate/recognize. The procedure and instructions for the generate/recognize condition remained exactly the same as in Experiment 1. For the cued-recall condition, subjects were informed that some stems in the test list could not be completed with an earlier presented word and were cautioned not to report words that were not presented earlier. Subjects were given feedback in the form of a loud beep when they mistakenly reported a new word as old. Otherwise, the procedure for the cued-recall test condition was the same as in Experiment 1.

Results and Discussion

The probability of solving anagrams in the first part of the experiment was .87. Results for the test phase will be reported separately for word stems allowing single and multiple completions.

Single Completions

Words earlier presented as anagrams were much more likely to be correctly recognized than were words that were earlier read (.84 vs. .54), $F(1,23) = 156.91$, $MSe = .007$. The probabilities of generating a word that was earlier presented as an anagram, that was earlier read, and that was new were .85, .83, and .65, respectively.

The probability of false recall in the cued-recall condition did not differ significantly from the probability of generating and falsely recognizing a new word as old (.11 vs. .09), $F < 1$. Examining performance on stems that could be completed with an old word (see Table 1), the probability of responding with an old word for cued recall was slightly higher than was the probability of generation and recognition (.68 vs. .59), $F(1,46) = 4.84$, $MSe = .041$. For both test conditions, words earlier presented as anagrams were more likely to be treated as old than were words earlier read (.76 vs. .51), $F(1,46) = 96.41$, $MSe = .015$. The interaction between anagram vs. read and test condition (see Table 1) did not approach significance, $F < 1$. This lack of a significant interaction contrasts with the results of Experiment 1 and can be taken as evidence that the increased reliance of cued recall on recognition processes, produced by the change in procedures, increased the similarity between processes underlying cued-recall performance and those underlying performance in the generate/recognize condition.

Results from the cued-recall condition in Experiment 2 were compared with those from Experiment 1 (see Table 1). Comparisons of results across experiments are always somewhat risky. However, the differences between the two experiments, aside from those in the procedure for the cued-recall test, at most, would be expected to produce only a significant main effect, not a significant interaction between conditions of prior presentation and experiment. In particular, although there was a slight difference between experiments in the ease of completing stems, there is no reason to expect that slight difference would influence the advantage in cued-recall performance of words presented as anagrams over those that were earlier read.

In Experiment 1, the probability of false recall was much higher than it was in Experiment 2 (.46 vs. .11), $F(1,46) = 55.73$, $MSe = .026$. Similarly, when stems could be completed with an old word, the overall probability of cued recall was higher in Experiment 1 than in Experiment 2 (.80 vs. .68), $F(1,46) = 9.48$, $MSe = .033$. The main effect of read vs. anagram was also significant, $F(1,46) = 24.75$, $MSe = .019$. More important, the interaction of read vs. anagram with experiment was significant, $F(1,46) = 9.64$, $MSe = .019$. As expected, the heavier reliance on recognition processes in Experiment 2 compared with Experiment 1 produced a much larger decrease in the probability of cued recall for words earlier read (.77 vs. .56) than it did for words earlier presented as anagrams (.82 vs. .79). That is, the change in procedures brought performance in the cued-recall condition into line with that in the generate/recognize test condition.

Multiple Completions

Words presented as anagrams were more likely to be recognized later than were words that were read (.70 vs. .43), $F(1,23) = 20.48$, $MSe = .045$. In the generate/recognize condition, the probabilities of generation for anagram, read, and new words were: .37, .41, and .90 (.28), respectively.

The pattern of results observed for multiple-completion word stems (see Table 3) was the same as that observed for single-completion word stems. The probability of false recall in the cued-recall condition did not differ significantly from the probability of generating and falsely recognizing a new word (.09 vs. .10), $F < 1$. For old words, the probability of cued recall was slightly higher than the probability of generation and correct recognition (.30 vs. .23), $F(1,46) = 4.81$, $MSe = .021$. In both test conditions, words earlier presented as anagrams were more likely to be treated as old than were words earlier read (.31 vs. .22), $F(1,46) = 21.11$, $MSe = .009$. The interaction between read vs. anagram and test condition (see Table 3) did not approach significance, $F < 1$.

Comparing results across experiments, the probability of false recall was much higher in Experiment 1 than it was in Experiment 2 (.48 vs. .09), $F(1,46) = 33.79$, $MSe = .056$. In both experiments, words earlier presented as anagrams were more likely to be recalled later than were words earlier read (.35 vs. .29), $F(1,46) = 6.63$, $MSe = .012$. Although the advantage of solving anagrams over reading words was larger in Experiment 2 than it was in Experiment 1, the relevant interaction was not significant, $F(1,46) = 1.37$.

The change in procedures had the expected effect of making cued recall more heavily reliant on recognition processes. Performance in the cued-recall condition in Experiment 2 was generally very similar to that in the generate/recognize condition. However, one remaining difference between performance in the two test conditions was that the probability of cued recall was higher than the probability of generation and recognition. This was true even when only a single completion of a word stem was possible and the probability of false recall did not differ significantly from that of generating and falsely recognizing a new test word. That pattern of results might be taken as evidence that cued-recall instructions compared with stem-completion instructions influenced generation processes, rather than serving only to add recognition processes to unchanged generation processes.

However, the difference between the probability of cued recall and that of generating and recognizing an old test word can be explained in terms of differential involvement of recognition processes. When a word came very promptly to mind as a completion for a word stem, subjects in the cued-recall condition may not always have done a recognition check to qualify the word for output as recalled. This might be true even though the probability of false recall in the cued-recall condition was not significantly higher than that in the generate/recognize condition. If a sufficiently high criterion value of fluency was used to qualify words for cued recall, nearly all of the words that passed that criterion would be words that were earlier presented. This means that a failure to do a recognition check would not necessarily carry the cost of producing a significant increase in the probability of false recall. Subjects in the generate/recognize condition were required to make recognition decisions for all generated words, regardless of how promptly a generated word came to mind. Recognition failure for old words that passed the fluency criterion for cued recall would have the effect of producing a probability of generation and recognition that was lower than the probability of cued recall. In line with the possibility that differential reliance on recognition processes was not fully removed, the advantage of anagram over read words was still slightly larger, although not significantly so, in the generate/recognize than it was in the cued-recall test condition.

For multiple-completion stems, subjects given cued-recall instructions were free to continue generating completions as candidates for recall until they recognized one as old. Doing so would have the effect of increasing the probability of responding with an old word. In contrast, subjects in the generate/recognize test condition were only allowed to generate one completion for each word stem. Consequently, when stems allowed multiple completions, it is surprising that the probability of responding with an old word for cued recall was not a great deal higher than the probability of generating and recognizing an old word. That similarity in performance, across test conditions, for multiple-completion stems might have resulted from the presence of single-completion stems in the test list. Because the generation of more than one completion was impossible for many stems in the test list, subjects may not have attempted to generate multiple completions, even when doing so was possible.

General Discussion

Tulving (e.g., Tulving & Thomson, 1973) used the finding of recognition failure for recallable words to dismiss generate/recognize models of recall. However, as well as recognition failure for recallable words, we found false recall of words that were correctly rejected on a test of recognition memory. In Experiment 1, the probability of cued recall for old words was sometimes higher than the probability of recognizing those old words in a generate/recognize condition, paralleling results reported by Tulving and Thomson (1973). The probability of false recall in the cued-recall condition was also much higher than the probability of generating and falsely recognizing a new word as old. The full pattern of results can be interpreted in terms of a generate/recognize model of recall that postulates two bases for memory decisions. If the generation of a response is sufficiently fluent, it is not subjected to a recognition check before being output as recalled. Interactions between type of test and other variables can also be explained in terms of that generate/recognize model. For example, the interaction between form of presentation (read vs. anagram) and stem completion (an indirect test of memory) vs. cued recall (a direct test of memory) is understandable if cued-recall but not stem-completion performance relies on recognition processes. As shown by comparing the results of Experiments 1 and 2, the reliance of cued recall on recognition processes can be increased by encouraging subjects to avoid falsely recalling new words as old.

As noted by Watkins and Gardiner (1979), generate/recognize models of recall, as originally formulated, are consistent with Tulving's (1985) distinction between episodic and semantic memory. By the structural assumptions of those models (e.g., Kintsch, 1970), there exists an abstract, semantic-memory representation of items, and a particular occurrence of an item is remembered by attaching a "tag" to its corresponding abstract representation. Attacks on those structural assumptions can as well be treated as attacks on the semantic/episodic distinction as attacks on generate/recognize models of recall. For example, a finding of effects of encoding specificity on perceptual-identification or stem-completion performance amounts to showing an effect of an episodic-memory variable on a semantic-memory task (Allen & Jacoby, in press; Jacoby, 1983; Roediger & Blaxton, 1987), and calls into question the claim that there is a semantic-memory representation that is separate from episodic memory for the earlier presentation of an item. Results of that sort should discourage attempts to describe differences between performance on indirect vs. direct tests of memory in terms of separate memory systems (e.g., Cohen & Squire, 1980; Tulving, 1985). Specifically, claims that semantic (or procedural) memory is used for indirect tests, but episodic memory is used for direct tests of memory are made less convincing when it is shown that memory for prior episodes influences performance on both types of tests. The finding that effects on performance of an indirect test of memory are specific to the prior processing of an item (e.g., read vs. anagram) is contradictory to claims (e.g., Graf & Mandler, 1984) that the activation of an abstract representation mediates those effects.

The generation process is influenced by memory for particular prior episodes, but those effects do not specify their source (e.g., Jacoby et al., 1989a). The fluency of generating an item in response to a question about memory is influenced by factors in addition to memory for a particular prior episode and, so, memory decisions based on judgments of fluency will sometimes be in error. The false recall observed in the present experiments likely resulted from the use of fluency as a basis for mistakenly judging generated words as remembered. In other experiments, it has been shown that misattributions of fluency can make nonfamous names seem famous (e.g., Jacoby, Woloshyn & Kelley, 1989b); influence subjective experience of physical parameters such as the loudness of a background noise (Jacoby, Allan, Collins, & Larwill, 1988); produce false recognition (Jacoby & Whitehouse, 1989); and lead to errors in the prediction of the performance of others (Jacoby & Kelley, 1987).

Models of the same general form as the model we adopt to account for differences between effects on direct and indirect tests of memory have been used to explain performance on a variety of other tasks, including perceptual (Broadbent, 1977); categorization (e.g., Rips, Shoben, & Smith, 1973); and question-answering tasks (Glucksberg & McCloskey, 1981; Reder, 1987). A generate/recognize model also seems particularly well-suited to explain differences in memory monitoring (e.g., Johnson & Raye, 1981) and differences in monitoring in social situations (e.g., Snyder, 1974). In that vein, Moscovitch (1989) has described the confabulations produced by amnesics as reflecting a deficit in editing or recognition processes. The difference between reconstruction and reproduction (e.g., Hasher & Griffin, 1978) can also be described in terms of a generate/recognize model. Hayman and Tulving (1989) found a correlation between recognition and cued-recall performance that was higher than the correlation between recognition and fragment-completion performance. They interpreted that difference in correlations as reflecting the differential involvement of a QM memory system. Their QM memory system might correspond to the recognition processes in generate/recognize models of recall.

The generate/recognize model that we propose is similar to a model proposed by Humphreys, Bain, and Pike (1989). The primary differences between our model and theirs are in the forms of knowledge said to underlie generation processes and in our claim that the reliance of recall on recognition-memory processes varies across situations. However, the similarities are probably more important than are the differences. For example, we agree with Humphreys et al. (1989) that the existence of separate memory systems is not necessary for differences between performance on indirect and direct tests of memory to be observed.

A model of the sort that we propose increases the complexity of predicting interactions between indirect vs. direct tests of memory and other factors. A variable relation between performance on the different types of tests is expected because neither direct nor indirect tests are "pure" with regard to the processes that they involve (e.g., Jacoby & Brooks, 1984). Manipulations of materials and of processing that one might expect to produce interactions with type of test do not always do so. For example, Weldon et al. (1989) presented items as words or as pictures and failed to find a significant interaction between mode of presentation and the probability of fragment completion vs. recall cued with word fragments. We would expect an interaction because words vs. pictures, like read vs. anagram, should have an effect on generation processes (fragment-completion performance) that is opposite to the effect it has on recognition-memory processes. However, predicting interactions requires that one look at the probability of false recall to assess the reliance of recall on recognition processes (Weldon et al. did this), and also that one have some estimate of the probability of recognition memory. The finding of interactions depends on the balance of effects on generation and effects on recognition-memory processes.

Predictions are even more difficult for effects on performance of complex tasks, such as recall of prose, than they are for recall cued with word fragments or word stems. In the latter case, cued-recall vs. completion instructions have little, if any, effect on generation processes, but, rather, have their effect by influencing recognition-memory processes. In contrast, when memory for prose or autobiographical memory is tested, test instructions almost certainly influence both generation and recognition-memory processes. Given recall instructions, subjects are likely to integrate information from the instructions with other cues (e.g., Humphreys et al., 1989) or may engage in reconstructive memory activities that serve to elaborate the cues provided for recall (e.g., Baddeley, 1982). What is needed are procedures that can be used to separate those effects on generation processes from effects on recognition-memory processes. The development of such procedures is the goal of research that we are currently doing.

Let us end by very briefly describing the rationale that underlies our current research. We started our current research by noting that the task of separating effects on generation processes from those on recognition-memory processes is made difficult by the standard procedure of arranging the test situation such that effects of the two types operate in the same direction. For example, cued recall can be enhanced either by increasing the ease of generation or by increasing the ease of recognition memory, making it difficult to specify the source of beneficial effects on cued-recall performance. As earlier argued for the more general problem of separating conscious from unconscious influences of memory (e.g., Jacoby et al., 1989b), advantages can be gained by arranging the situation such that effects on generation and effects on recognition-memory processes are in opposition to one another. Applying an opposition strategy, one might instruct subjects to complete word stems with a word that was *not* presented earlier. Earlier presenting a word that could be used to complete a stem would then enhance generation of that word as a completion. However, that effect on generation processes would be opposed by recognition-memory processes, because subjects would withhold the word as a response if they recognized it as old. In that circumstance, for prior presentation to increase the probability of a word being given as a completion, effects on generation processes must outweigh effects on recognition-memory processes.

We are developing procedures of the above sort to more clearly separate effects on generation from those on recognition-memory processes, using a variety of different materials. It is important to develop new procedures to separate effects on the two types of processes, rather than abandoning generate/recognize models of recall. Generate/recognize models are too useful as descriptions of memory monitoring and other activities to be abandoned.

Appendix A: Stimuli for Experiment 1

Low-frequency words completions				Medium-frequency words completions			
Multiple		Single		Multiple		Single	
word	anagram	word	anagram	word	anagram	word	anagram
clamp	plcma	aisle	eisla	brick	krbci	agent	egant
clown	nlcwo	bylaw	wybal	bleed	eldeb	apron	rpnoa
dowry	yodrw	crypt	trypc	chalk	ahclk	bacon	canob
drawl	ardwl	deuce	ueecd	cliff	flifc	essay	ssyae
prism	mrisp	farce	eafcr	drill	lrdli	fable	eablf
patio	taoip	icing	gcini	flush	ulhsf	lodge	eodgl
plaza	alpza	irony	oryni	mercy	reycm	nerve	eenvr
spasm	apmss	oasis	sasio	stalk	atslk	piano	aipno
spurt	uptrs	pedal	delap	spear	rpeas	punch	nupch
trash	artsh	thyme	ehtmy	thief	ihtef	token	koten
broth	hrbto	album	blmua	basis	sabis	agony	ogany
cleat	tleac	bigot	tibog	bloom	mlboo	acorn	ncora
crock	orkcc	dunce	eudcn	blush	ulhsb	feast	teasf
chili	ihilc	easel	salee	craft	trafc	geese	eegse
forum	mofur	fudge	eufgd	creek	erkec	ivory	yvori
greed	drgee	jaunt	uajnt	growl	lrgwo	novel	volen
gland	algnd	pooch	oopch	scarf	acsrf	opera	eparo
mania	aamin	pygmy	yygmp	stump	ptsmu	ounce	euocn
probe	erpbo	rayon	naroy	spice	epics	pulse	eupsl
slime	ilems	yacht	tachy	twine	ewint	upset	spuet
boost		algae		crust		bench	
chess		bison		globe		index	
froth		cynic		strap		ledge	
floss		ebony		spoon		olive	
snack		tooth		truck		organ	

References

Allen, S. W., & Jacoby, L. L. (in press). Reinstating study context produces unconscious influences of memory. *Memory and Cognition.*

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review,* **79,** 97–123.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology, Vol. 1, Learning, memory and thinking* (pp. 243–293). San Francisco: Freeman.

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review,* **89,** 708–729.

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review,* **77,** 215–222.

Bartlett, F. (1932). *Remembering.* Cambridge: Cambridge University Press.

Broadbent, D. (1977). The hidden preattentive processes. *American Psychologist,* **32,** 109–118.

Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Categories reconsidered: The ecological and intellectual bases of categories* (pp. 141–174). Cambridge: Cambridge University Press.

Cofer, C. N., Chmielewski, D. L., & Brockway, J. F. (1976). Constructive processes and the structure of human memory. In C. N. Cofer (Ed.), *The structure of human memory.* San Francisco: Freeman.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science,* **210,** 207–210.

Dooling, D. J., & Christiaansen, R. E. (1977). Episodic and semantic aspects of memory for prose. *Journal of Experimental Psychology: Human Learning and Memory,* **3,** 428–436.

Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory,* **7,** 311–325.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior,* **23,** 553–568.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **10,** 164–178.

Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and Recognition* (pp. 183–216). London: Wiley.

Hasher, L., & Griffin, M. (1978). Reconstructive and reproductive processes in memory. *Journal of Experimental Psychology: Human Learning and Memory,* **4,** 318–330.

Hayman, C. A. G., & Tulving, E. (1989). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory & Cognition,* **15,** 228–240.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychology Review,* **93,** 411–428.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review,* **96,** 208–233.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior,* **17,** 649–667.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior,* **22,** 485–508.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **14,** 240–247.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 1–47). New York: Academic Press.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General,* **3,** 306–340.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin,* **13,** 314–336.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989a). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General,* **118,** 126–135.

Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989b). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General,* **118,** 115–125.

JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review,* **88,** 67–85.

KELLEY, C. M., JACOBY, L. L., & HOLLINGSHEAD, A. (1989). Direct versus indirect tests of memory for source: Judgments of modality. *Journal of Experiment Psychology: Learning, Memory and Cognition,* **15,** 1101–1108.

KINTSCH, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 333–370). New York: Academic Press.

KINTSCH, W. (1978). More on recognition failure of recallable words: Implications for generation-recognition models. *Psychological Review,* **85,** 470–473.

MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review,* **87,** 252–271.

MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review,* **85,** 207–238.

MOSCOVITCH, M. (1989). Confabulation and the frontal systems: Strategic versus associative retrieval in neuropsychological theories of memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 133–156). Hillsdale, NJ: Erlbaum.

NELSON, D. L., CANAS, J. J., BAJO, M. T., & KEELEAN, P. D. (1987). Comparing word-fragment completion and cued recall with letter cues. *Journal of Experimental Psychology: Learning, Memory & Cognition,* **13,** 542–552.

NELSON, D. L., KEELEAN, P. D., & NEGRAO, M. (1989). Word-fragment cuing: The lexical search hypothesis. *Journal of Experimental Psychology: Learning, Memory & Cognition,* **15,** 388–397.

REDER, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology,* **19,** 90–138.

RICHARDSON-KLAVEHN, A., & BJORK, R. A. (1988). Measures of memory. *Annual Review of Psychology,* **39,** 475–543.

RIPS, L. J., SHOBEN, E. J., & SMITH, E. F. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior,* **12,** 1–20.

ROEDIGER, H. L., & BLAXTON, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and cognitive processes: The Ebbinghaus centennial conference.* Hillsdale, NJ: Erlbaum.

SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory,* **4,** 592–604.

SNYDER, M. (1974). The self-monitoring of expressive behavior. *Journal of Personality & Social Psychology,* **30,** 526–537.

THOMSON, D. M., & TULVING, E. (1970). Associative encoding and retrieval: weak and strong cues. *Journal of Experimental Psychology,* **86,** 255–262.

THORNDIKE, E. L., & LORGE, I. (1944). *The teachers' handbook of 30,000 words.* New York: Columbia University Press.

THORNDYKE, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive psychology,* **9,** 77–110.

TULVING, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 37–73). London: Wiley.

TULVING, E. (1985). How many memory systems are there? *American Psychologist,* **40,** 385–398.

TULVING, E., & THOMSON, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review,* **80,** 352–373.

WATKINS, M. J., & GARDINER, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning and Verbal Behavior,* **18,** 687–704.

WELDON, M. S., ROEDIGER, H. L., & CHALLIS, B. H. (1989). The properties of retrieval cues constrain the picture superiority effect. *Memory & Cognition,* **17,** 95–105.

(Received August 4, 1989)

(Revision received September 21, 1989)