Memory & Cognition
1990, 18 (1), 72-82

Effects of repetition of mental operations on memory for occurrence and origin

JAN C. RABINOWITZ
Barnard College, Columbia University, New York, New York

In two experiments, subjects read or generated items at both encoding and retrieval. At test, they were required to decide whether or not the targets were presented initially (recognition), and if so, whether they were initially read or generated (judgments of origin). Recognition for items that were initially generated was enhanced if they were once again generated at test in the same context, but not if they were generated at test without context. These results confirm that memory for occurrence is facilitated by repetition of the initial encoding operations at retrieval. Generating at test resulted in an increase in "generate" responses both for items that were initially generated and for items that were initially read. Overall, there was a decrease in the accuracy of origin discriminations. It is suggested that, when subjects generate at test, they are likely to mistakenly attribute these just-performed operations to be part of the memory trace for that item.

Understanding the cognitive operations, or mental processes, by which we perceive and imagine events is a central focus of cognitive psychology. More recently, it has been argued that repetition of cognitive operations is a critical determinant, or perhaps, the determinant of remembering (Kolers, 1975, 1979; Kolers & Roediger, 1984). The focus of this paper is to examine further the effect of repetition of cognitive operations on memory for occurrence and to begin to explore the effects of repetition of cognitive operations on memory for the origin, or source, of a remembered event.

Glisky and Rabinowitz (1985) reported that repetition of the same cognitive operations during retrieval as those that were performed during encoding resulted in an increase in recognition. We required subjects to read target words or generate words from word fragments (e.g., AL-OHO-). At test, they were again required to read or generate the words before making recognition decisions. In addition to the standard generation effect of better memory for the items presented as fragments (Slamecka & Graf, 1978), we observed what we called a *specific generation enhancement* effect. For those items that were initially generated, recognition performance was improved if they were generated again, rather than read, at test. There was no effect of reading or generating at test for those items that were initially read.

We therefore argued that a complete account of the generation effect must include a specific processing component. Records of cognitive operations are stored as part of the memory trace of the stimulus event (Johnson & Raye, 1981; Kolers, 1975, 1979; Russo & Wisher, 1976). Repeating the same operations at test as those that were performed at encoding facilitates access to the memory trace of which those operations are a part. This view is supported by two additional findings. First, a generation enhancement effect was not observed when the letters that were missing from the target at retrieval were different from the letters missing at encoding (AL-OH-L). Thus, the generation enhancement effect is not simply due to regenerating the same word. Rather, the same generation operations must be repeated. Second, a generation enhancement effect was not observed even if the same letters were missing from the targets at retrieval, when the generation process could be guided by a different type of information. We had subjects read or generate single target words at encoding. In one test condition, the targets were tested along with an extralist semantically related cue (booze–AL-OHO-). We argued that this generation process, guided primarily by semantic relational information, would be quite different from the generation process used at encoding (when no cues were present), which would have been guided primarily by orthographic and phonemic information. Because the generation operations were different, no benefit from generating at test was predicted, and none was observed.

One purpose of the present experiments is to replicate the specific generation enhancement effect using different generation operations and different materials. Both experiments required subjects to read or generate target items at both encoding and retrieval. It is predicted that for items that are initially generated, reinstating the same generation operations at test will result in an additional recognition benefit.

This research was supported by National Institute of Mental Health Grant MH42665. Portions of this research were presented at the annual meeting of the Psychonomic Society, Seattle, November 1987. I thank Risa Evans, Karen Dine, Eva Pomerantz, and Ting Lan Chin for experimental assistance. Thanks also to Fergus Craik, Helene Intraub, Larry Jacoby, David Elmes, David Payne, and especially, Marcia Johnson, for comments on earlier versions of this manuscript. Reprint requests should be sent to Jan Rabinowitz, Department of Psychology, Barnard College, 3009 Broadway, New York, NY 10027.

—Accepted by previous editor, Alice F. Healy

The second major focus of this paper is to examine the effect of repetition of cognitive operations on memory for origin, that is, whether the event was originally read or generated. Johnson and Raye (1981) argued that remembering information about previously completed cognitive operations provides a critical cue for determining whether an event was originally presented externally or was internally imagined. They hypothesized that externally presented events are often perceived relatively automatically, whereas internally generated thoughts result in more information regarding the cognitive operations used in producing the memory. Therefore, if the cognitive operations of the event are remembered, the event is more likely to have been internally generated.

Much of the evidence for the role of information regarding cognitive operations in judgments of origin (reality monitoring) has been obtained from generation paradigms. More difficult generation tasks are hypothesized to produce more distinctive records of cognitive operations, which are more easily accessed (Johnson, Raye, Foley, & Foley, 1981). For example, the difficulty of the generation task can be manipulated by varying the taxonomic frequency of the targets that subjects must generate in response to a category cue. In experiments of this sort, judgments of origin for generated low-taxonomic-frequency items are more accurate than judgments of origin for generated high-taxonomic-frequency items. There is no effect of taxonomic frequency on the accuracy of origin judgments for items that were initially read (Johnson et al., 1981; Rabinowitz, 1989).

Additional evidence for the role of cognitive operations in judgments of origin comes from tasks in which the number of encoding presentations of read and generate items is manipulated (Johnson, Raye, & Durso, 1980; Rabinowitz, 1989). Subjects are more accurate at identifying the source of items that were generated twice at encoding than they are for items that were generated only once. However, there is no effect of the encoding repetition manipulation for items that were read. Thus, repeating generation operations at encoding increases the likelihood of subsequently remembering them. These data also suggest that it is the information about cognitive operations that weigh most heavily in origin decisions for whether words were read or generated. In fact, Johnson et al. (1981) have argued that, in generation tasks of this sort, the primary "discriminative cue that differentiates the classes of externally presented and self-generated memories is the greater amount of information about cognitive operations typically associated with self-generated traces" (p. 61).

Remembering the cognitive operations associated with a particular memory trace provides good evidence that the item was initially generated. The hypothesis to be tested is whether memory for those operations can be facilitated by repeating them at the time of test. Given a multicomponent view of the memory trace (e.g., Bower, 1967), it is quite possible to access the memory trace and recognize the item, but to be unable to access all of its component attributes. It is hypothesized that generating at retrieval should facilitate access to, and identification of, similar information that has been stored as part of the memory trace. In a sense, when an item is read at retrieval, subjects attempt to recall additional information about cognitive operations. When an item is generated at retrieval, the subjects now need to "recognize" whether these operations were performed before. To the degree that recognition of operations is easier than recall of operations, discrimination of origin should be more accurate when the operations are repeated at test. Intuitively, if an item was generated initially, and those same generation operations are repeated at test, subjects might better remember "doing that before."

For items that were initially read, no particular benefit for judging origin is to be expected from reading again at retrieval, just as is the case with item recognition. Because reading is such an automatic process, it has been suggested that records of this process are not stored, or at least are not discriminable from all the other occasions on which the word has been read (Glisky & Rabinowitz, 1985; Johnson & Raye, 1981). But when subjects generate at retrieval, a mismatch should result, because those operations were not stored as part of the memory trace at encoding, and the subjects should recognize that those operations were not performed before. Therefore, we may also predict that generating at retrieval should increase the accuracy of origin judgments of items that were initially read.

To summarize, these experiments were designed to determine whether or not repetition of the cognitive operations involved in completing word fragments facilitate access to the memory trace (recognition) and access to, or awareness of, the record of those operations as part of the trace in a manner that might facilitate judgments of origin. If identification of those operations as part of the memory trace is facilitated by repeating them at test, then it is predicted that generating at retrieval should increase the accuracy of origin judgments for items initially generated. The accuracy of origin judgments for items that were initially read should either be unchanged or also increased as a result of generating at retrieval. In either case, the net accuracy in discriminating whether an item was initially read or generated should be increased as a result of generating at retrieval.

EXPERIMENT 1

The first purpose of Experiment 1 was to replicate the specific generation enhancement effect, under what can be considered to be the reverse set of experimental conditions of those used by Glisky and Rabinowitz (1985) in their Experiment 5. In the present experiment, all subjects studied semantically related cue-target pairs, in which the targets were either presented intact (read) or as fragments (generate). For half of the subjects, the tar-

gets were tested in the presence of the original cue words (*context retrieval* condition). Once again, half of the targets had to be read and half had to be generated before subjects made their recognition decisions. The manipulations of read/generate at encoding and read/generate at retrieval were crossed factorially. A specific generation enhancement effect is predicted for these subjects. In addition to a main effect of generation at encoding, generate-generate items should be better recognized than generate-read items.

For the remaining subjects, all of the targets (and distractors) were tested without the original cue words (*no-context retrieval* condition). Half of the targets were read at test and half were generated. For these subjects, the generation operations at test, when no cue is present, should be guided primarily by orthographic and phonemic information. The semantic component, which was present at encoding, is no longer present. Thus, there should be relatively little overlap between the generation operations used at encoding and retrieval. Therefore, no effect of generating at test is predicted. Only a main effect of generation at encoding should be observed.

For each item judged "old" on the recognition test, the subjects were also required to decide whether the item was initially presented in read or generate form. If the identification or recognition of cognitive operations as a part of the memory trace can be facilitated by performing them at the time of test, then the accuracy of the origin judgments should increase when subjects are required to generate at test. This prediction applies only to those situations when the generation operations are the same at study and test (the context retrieval condition). When the generation operations at test are different from those at study (the no-context retrieval condition), no benefit in the accuracy of the origin judgments should be expected from generating at test.

The decision to obtain origin judgments only for those items that are recognized was made, knowing that it subjects these data to potential confounds from item selection effects, given that the recognition hit rates are expected to differ among conditions. But, consider an alternative experimental design that might attempt to avoid these item selection effects. Subjects could have been tested with only the old items and required to make an origin decision for each item. There are two problems with this procedure. First, even though subjects are told that all of the items are old, they will not remember some of them. From the subject's perspective, it is not very meaningful to judge how an event initially occurred, when its occurrence itself is not remembered. Second, this procedure introduces a bias to the data that is even more severe than the potential ones we might want to eliminate. How will subjects respond to those events that are not remembered? One reasonable possibility is that, for these events, the accuracy of the origin judgments will be close to chance, or at least less than the accuracy for other events from this class that are remembered. For any given experimental class of events, the overall accuracy of the origin judgments will then be a function of the accuracy of the judgments for the remembered events, the nonremembered events, and the relative proportions of these two types of events. This would result in a systematic bias such that the accuracy of the origin judgments for every class of events will be reduced by an amount that is inversely related to the level of recognition accuracy for that class of events. This would tend to induce a positive correlation between levels of recognition performance and accuracy of origin judgments across experimental conditions.

Alternatively, subjects may simply respond "read" to all nonremembered events. This is also a reasonable strategy, especially if subjects realize that read items are less likely to be remembered than generated items. Under this assumption, all origin judgments for nonremembered read items would be correct and all origin judgments for nonremembered generate items would be incorrect. This would induce a negative correlation between levels of recognition performance and accuracy of origin judgments across experimental conditions for different classes of items that were initially read, and a positive correlation for items that were initially generated. In addition, the accuracy of the origin judgments for items initially read would be increased relative to the items that were initially generated.

It is of both empirical and theoretical interest to determine the factors that affect origin judgments for nonremembered items. However, the above analyses have shown that, given either of two reasonable models for these origin judgments, experimental designs that require subjects to make origin decisions for all old items would result in the introduction of systematic biases to the origin data.

Furthermore, and perhaps more importantly, the phenomenon under study is how people determine the origin of remembered events. If classes of events differ in their memorability, then memory for the origin of these remembered events must, by definition, include different numbers of instances from each class. Item selection effects are not a problem in the data; they are an inherent part of the problem under study. The fact that events differ in their overall memorability cannot preclude our asking questions about memory for their origins.

Method

Subjects. Eighty subjects participated in Experiment 1; half were randomly assigned to the context retrieval condition and half to the no-context retrieval condition. Sixteen of the subjects in each group were respondents to advertisements at Columbia University asking for paid subjects for a memory experiment. They received $4 for participating in the experiment. The remaining subjects in each group were volunteers from introductory psychology classes at Barnard College. They received bonus course points in return for their participation.

Materials. Two sets of 100 cue-target pairs were created. Each cue was associatively related to two target words (SCHOOL-

STUDENT and SCHOOL-TEACHER). A generate version of each target word was formed by replacing two letters with hyphens (SCHOOL-ST-DE-T). The first letter of the target words was never omitted.

Design. The experiment was a 2 × 2 × 2 mixed factorial design. The within-subject variables were read/generate at encoding and read/generate at retrieval. The between-subject variable was the presence or absence of context (the encoding cue word) at retrieval.

The presentation forms contained 100 cue-target pairs. On each presentation form, 50 of the targets appeared in the read condition and 50 appeared in the generate condition. All targets were presented with their associated cue words. The order of the cue-target pairs on each presentation form was random. Four different presentation forms were used: two contained the first set of target words and two contained the second set of target words. For each set of target words, the two presentation forms differed in that those targets that were read on the first form were generated on the second form, and vice versa.

The test forms contained both the set of targets that appeared on the presentation form and the other 100 target words as distractors. Next to each test item were printed the words "New/Old" and "Read/Generate." Two test forms were prepared for the context test conditions and two for the no-context test conditions. On each form, half of the items were read and half were generated; the assignment of items to test condition was reversed across the two forms. The forms were prepared so that, on a given form, targets were tested in the same test condition as their matched distractors. Thus, if the target word STUDENT was tested in generate form, then the matched distractor TEACHER was also tested in generate form. This ensured that there were 25 targets in each of the four conditions formed by the factorial combination of read/generate at encoding × read/generate at retrieval, as well as 50 read distractors and 50 generate distractors. The order of the 200 items on the test form was random. The two test forms for each test condition were crossed with the four presentation forms such that items appeared equally often as targets and distractors, and each item appeared in each test condition equally often across subjects.

Procedure. The subjects were instructed to associate each target with its cue and were told that their memory for the target words would be tested. The nature of the test was not specified. For generate items, they were instructed to use the cues to help them determine the missing letters for each target and to write in the missing letters. The subjects worked with a cardboard mask that allowed only one cue-target pair to show at a time; they moved the mask at 4-sec intervals, which were indicated by a prerecorded set of brief tones. The subjects were told to skip any words that they could not generate within the interval indicated by the timer.

After the presentation, the subjects were given a distractor task in the form of a short vocabulary test. They were then given the memory test. They were instructed to read or generate each target word on the test as they came to it and to indicate whether it was *new* (i.e., not seen on the presentation form) or *old* (i.e., seen as a target word on the presentation form). For items judged "old," they were asked to indicate whether they had read the item or generated it on the presentation form. Recognition judgments and judgments of origin were made for the single items in the no-context retrieval condition and for the target items in the context retrieval condition. The subjects in the context retrieval condition were told that all of the cue words had appeared on the presentation form and that the target words were always matched with the same cues as they had been matched with on the presentation form. Therefore, their recognition and judgment of origin decisions were to be made for the second word of each pair, and they were told that the cue was present to help them make their decisions. All of the subjects were also informed that the breakdown of new and old items was roughly half and half. They were instructed to complete all decisions about one item before moving on to the next and not to backtrack. Each subject used the mask, and worked at his/her own pace.

Results

Items were scored as incorrectly generated if the subject failed to generate the correct word during presentation and/or test. Overall, the proportions of incorrectly generated items were quite low, but they did differ as a function of study-test condition and subject group. The mean proportions of incorrectly generated items in the context test group were .039, .006, and .028, for the generate-read, read-generate, and generate-generate conditions, respectively. (The value for the generate-generate condition represents the proportion of items that were incorrectly generated at study and/or test.) The comparable values for the no-context test group were .034, .033, and .033. There was a significant main effect of item type [$F(2,78) = 5.62$, $MS_e = 0.001$], and a significant interaction between context and item type [$F(2,78) = 5.06$]. (Alpha was set to .05 for all analyses.) Both of these effects can be attributed to the difference for the read-generate items. Items initially read were almost never incorrectly generated on the test form when they were tested with context. However, if they were tested without context, they did not benefit from having been read initially. The context test and no-context test groups also differed in the proportion of distractor items on the recognition test that were incorrectly generated [$F(1,78) = 17.11$, $MS_e = 0.002$]. The means were .022 for the context test group and .064 for the no-context test group, reflecting the fact that the presence of semantic relational information resulted in an easier generation task.

Memory for occurrence. Only the data for correctly generated items were analyzed. For old items, an item was omitted from the calculation of the hit rates if it was incorrectly generated at either study or test. For the new items, the false-alarm rates for the generate test items were based on only the items that were generated correctly at test.

The false-alarm rate was higher in the no-context test condition (.24) than in the context test condition (.17) [$F(1,78) = 7.91$, $MS_e = 0.025$]. The false-alarm rate was also higher for items that were generated at test than for items that were read at test [$F(1,78) = 9.08$, $MS_e = 0.006$]. This difference was slightly larger in the no-context test condition (.05) than in the context test condition (.02), but the interaction between these two factors was not significant [$F(1,78) = 1.07$].

Because of the false-alarm-rate differences, d' values were computed for each subject in each of the four conditions. These values are shown in Table 1, along with the hit rates and false-alarm rates. All of the statistical analyses were performed on the d' values. The first major set of predictions involved the generation specificity effect: Are items generated at study better recognized when they are generated again at test than when they are read at test? This benefit was present in the context retrieval

Table 1
Mean Recognition Hit Rates, False-Alarm Rates, and *d'* Values for Experiment 1

	Encoding Condition				
	Read		Generate		False-Alarm
Retrieval	Hit Rate	*d'*	Hit Rate	*d'*	Rate
		Context Test			
Read	.67	1.50	.77	1.80	.16
Generate	.67	1.45	.85	2.02	.18
		No-Context Test			
Read	.64	1.21	.82	1.78	.22
Generate	.66	1.13	.86	1.78	.27

condition (a difference of .22 *d'* units), but not at all in the no-context retrieval condition.

Separate analyses were done for the context and no-context groups. In the context retrieval condition there was a significant overall advantage for generating at encoding [$F(1,39) = 31.52$, $MS_e = 0.241$], and a significant interaction between read/generate at encoding and read/generate at retrieval [$F(1,39) = 6.24$, $MS_e = 0.110$]. The main effect of generating at retrieval was not significant [$F(1,39) = 1.16$, $MS_e = 0.254$]. A priori contrasts confirmed that the advantage of the generate-generate condition over the generate-read condition was significant [$t(39) = 2.32$], but that the slight advantage of read-read over read-generate was not.

In the no-context retrieval condition there was also a significant main effect of generating at encoding [$F(1,39) = 111.75$, $MS_e = 0.133$]. However, generating at retrieval did not result in an advantage for items generated at encoding. The main effect of retrieval condition was not significant [$F(1,39) = 0.36$, $MS_e = 0.139$], nor was the encoding × retrieval interaction [$F(1,39) = 0.77$, $MS_e = 0.083$].

This experimental design also allows us to assess the degree to which relational information is strengthened by generation at encoding. It is well known that when target items are studied in the presence of related cue words, recognition of the targets is better if they are tested in the presence of the original cue words than if they are tested alone (Thomson, 1972; Tulving & Thomson, 1971). This is because recognition is based, in part, on a retrieval process in which access to the entire pair is attempted (Mandler, Rabinowitz, & Simon, 1981). The provision of the cue (context) at the time of test improves recognition by reinstating the semantic relational information between the words and providing access to the entire stored pair. Should a similar recognition benefit from testing with context be observed if the targets are generated, rather than read, at encoding?

Rabinowitz and Craik (1986) suggested that the information enhanced by generation is the information that is actually used to guide the generation process. For example, when the generation process is guided by semantic relational information provided by a cue word, it is that same semantic relational information which is enhanced. If generation enhances relational information about the cue-target pair, then the accessibility of that same relational information and of the cue itself should be increased when the target is presented on the recognition test. Therefore, it was predicted that explicitly providing the cue at the time of test should provide a lesser benefit for targets that were originally generated.

To test this hypothesis, an additional ANOVA was performed for those items that were read at retrieval, with read/generate at encoding as a within-subject factor and context/no context at retrieval as a between-subject factor. As predicted, there was a significant interaction between these two factors [$F(1,78) = 5.32$, $MS_e = 0.134$]. The beneficial effect of context was present for items that were read at encoding [$t(78) = 2.08$], but not for items that were initially generated [$t(78) = 0.17$].

Memory for origin. Judgments of origin were obtained for only those items that the subjects recognized as old. The number of origin decisions thus varies widely across conditions and subjects, as a function of the hit rates. Therefore, an analysis was performed on the arcsine transformations of the proportions of correct judgments. The mean proportions of correct judgments of origin, shown in Table 2, were transformed back into proportions with a sine transformation.

For items that were initially generated, generating at retrieval resulted in an increase in the proportion of correct origin judgments. But this effect was not specific to repeating the same generation operations as were performed at encoding, since this increase occurred in both the context and no-context retrieval conditions. Furthermore, subjects were also more likely to respond "generate" to items that were initially read when they generated at retrieval. Thus, for these items, generating at retrieval resulted in a decrease in the proportion of correct origin judgments. Considering all the items, there was actually a decrease in the proportions of correct origin judgments when the subjects generated at retrieval.

The statistical analysis of these data confirmed these observations. Neither the main effect of context at retrieval nor any of the interactions involving this factor was significant. An overall bias towards responding "read" can be seen in the higher proportion of correct judgments for

Table 2
Mean Performance Levels for Judgments of Origin in Experiment 1

Retrieval	*p*("R"/R)*	*p*("G"/G)†	*d'*	*C*	*p*("R"/N)‡
		Context Test			
Read	.93	.66	1.65	−.45	.83
Generate	.80	.78	1.49	−.03	.74
		No-Context Test			
Read	.92	.70	1.69	−.37	.85
Generate	.84	.75	1.54	−.13	.78

*Probability of responding "read" to an item that was read at encoding, given that it was recognized. †Probability of responding "generate" to an item that was generated at encoding, given that it was recognized. ‡Probability of responding "read" to a new item, given that it was falsely recognized.

items initially read (.88) than for items initially generated (.72) [$F(1,78) = 51.11$, $MS_e = 0.114$]. Overall, the proportion of correct judgments of origin was higher when subjects read at retrieval (.82) than when they generated at retrieval (.79) [$F(1,78) = 8.90$, $MS_e = 0.023$]. There was also a significant interaction between an item's status (whether it was read or generated at encoding) and whether it was read or generated at retrieval [$F(1,78) = 35.55$, $MS_e = 0.065$]. Generating at retrieval resulted in a .08 increase in the proportion of correct "generate" judgments. But it also resulted in a .13 *decrease* in the proportion of correct "read" judgments. That is, subjects were more likely to respond "generate" to both read and generate items when they generated at retrieval.

Judgments of origin should be conceptualized as a discrimination task in which subjects must decide whether an item was originally presented in either read or generate form. A subject's judgment will be influenced both by his/her ability to discriminate and by any bias that he/she might have to respond "read" or "generate." Therefore, the judgments of origin were also analyzed using the methods of signal detection theory. For each subject, in each condition (read retrieval and generate retrieval), the proportion of read judgments for items that were initially read [p("R"/R)] was treated as a hit rate and the proportion of read judgments for items that were initially generated [p("R"/G)] was treated as a false-alarm rate. The latter value is simply the one's complement of the proportion of generate judgments for items that were initially generated [p("G"/G)]. From these values, d' and C (a measure of response bias) were calculated. Snodgrass and Corwin (1988) found that these measures of discrimination and response bias are independent (whereas d' and beta are not). C is defined as the distance of the response criterion from the intersection of the two underlying distributions. C ranges from -1 to $+1$, with 0 indicating a neutral response bias. Negative C values indicate a bias towards responding "read," and positive C values indicate a bias towards responding "generate." The mean d' and C values are also shown in Table 2.

First, there was no effect of the presence or absence of context at retrieval on either origin discrimination (d') or bias (C). None of the main effects or interactions involving this factor approached significance.

The read/generate retrieval manipulation had reliable effects on both d' [$F(1,78) = 6.04$, $MS_e = 0.154$] and C [$F(1,78) = 36.62$, $MS_e = 0.116$]. The effects of this retrieval manipulation were not as initially predicted but are consistent with what was observed in the analysis of the proportions of correct judgments. When the subjects read at retrieval, d' was greater than when they generated at retrieval. The subjects also exhibited a marked response bias towards responding "read" on the judgment of origin test. This bias was substantially reduced when they generated at test.[1]

Another measure of response bias was obtained from the judgments of origin for those new items called old (the false alarms). If no bias is present, the proportion of read judgments for new items should equal .50. The mean proportion of read judgments for new items was .84 in the read test condition and .76 in the generate test condition. Thus, a substantial read bias was present in all conditions, but it was reduced when the subjects generated at test [$F(1,78) = 6.27$, $MS_e = 0.111$]. The read bias for new items was not affected by the context manipulation.

Discussion

Memory for occurrence. Substantial generation effects were observed in all conditions of Experiment 1. Generation clearly requires more extensive processing of the target word than does reading. There appears to be a general consensus that this results in a greater activation of an item's semantic representation, resulting in a greater number of functional retrieval routes (Nairne, Pusen, & Widner, 1985).

The generation effect can also be modulated by the processes engaged at retrieval. When the same generation operations are repeated at test, the magnitude of the generation effect is increased. A recognition benefit is not observed from generating at test if the items were initially read, or if the generation operations at test are different from the ones at encoding. These results thus replicate and extend those of Glisky and Rabinowitz (1985) and attest to the role of repetition of mental operations in recognition (Cermak, Schnorr, & Buschke, 1970; Kolers, 1975, 1979; Russo & Wisher, 1976).

The presence of the original encoding context at retrieval resulted in a substantial recognition benefit for items that were initially read, thus replicating the findings of Tulving and Thomson (1971) and Thomson (1972). No benefit from context was observed for items that were generated at encoding, if they were read at retrieval. This absence of a benefit from context at retrieval for generated items has been replicated with a separate sample of 16 subjects. These subjects read or generated the same 100 cue–target pairs as in this experiment. All of the items were read at retrieval. Half of the items were tested with context, and half without context. The hit rates for the read items were .58 ($SD = .13$) in the no-context test condition and .67 ($SD = .13$) in the context test condition. The comparable values for the items that were initially generated were .77 ($SD = .14$) and .78 ($SD = .12$). (The false-alarm rates in the no-context and context test conditions did not differ.)

The absence of a benefit from reproviding the initial encoding context for initially generated items is consistent with the hypothesis that generation enhances the relational information used to guide the generation process (Donaldson & Bass, 1980; Hirshman & Bjork, 1988; Rabinowitz & Craik, 1986). A reasonable extension of this view would hold that retrieval, from the target, of this relational information and/or the cue word itself is increased for generated items. Thus, explicitly providing the cue on the test confers less of a benefit for generated items.

To the degree that retrieval of the relational information implies retrieval of the cue word, it would be predicted that generating the target in a cue-target pair should also increase access to the cue word from the target. This hypothesis was explicitly addressed by Slamecka and Graf (1978) in their pioneering explorations of the generation effect. They confined their investigation of this hypothesis to pairs of rhyming items. Absolutely no evidence of a generation effect was found when recognition of the cue (stimulus) was tested in the presence of the target (response) (Experiment 3). However, their data were less clear with regard to recall (Experiment 5). In a multitrial recall experiment, they reported a trend towards a larger generation effect for recall of the targets, given the cues, than for recall of the cues, given the targets. However, the interaction between these two factors was not significant. Indeed, recall on Trial 1 appeared to show equally large generation effects for stimulus and response recalls. They therefore prudently deferred drawing any firm conclusion as to whether or not generation also increases recall of the cue words, pending further replication. Recently, however, Greenwald and Johnson (1989) reported a generation effect for the recall of the stimulus terms of related pairs when they were cued with the response terms. Their results are thus consistent with our proposed explanation for the lack of a context effect for generated items.

Memory for origin. It was hypothesized that repetition of the cognitive operations involved in completing word fragments would facilitate the recognition or identification of those operations as part of the memory trace. Thus, it was predicted that generating at retrieval would increase the accuracy of judgments of origin. This effect was not observed. Generating at retrieval resulted in an increase in generate judgments for items that were initially generated (correct judgments) and for items that were initially read (incorrect judgments). On balance, there was actually a small, but significant, decrease in the ability to discriminate origin when subjects were required to generate at retrieval.

Before discussing the theoretical implications of these results, a replication was deemed necessary.

EXPERIMENT 2

Experiment 2 again required the subjects to read or generate items at both encoding and retrieval. At retrieval, the subjects were required to make recognition decisions for each item; for those items judged "old," the subjects were asked to decide whether the item was initially studied in read or generate form. The experimental materials consisted of category instances along with their category labels. Whenever the subjects were required to generate a category instance, the category label was provided along with the first three letters of an instance of the category.

Category instances of both high and medium taxonomic frequency were used in this experiment. This manipulation was included in order to provide a benchmark against which the judgment of origin results could be evaluated. Previous research (Johnson et al., 1981; Rabinowitz, 1989) has shown that subjects are more likely to correctly judge previously generated medium-taxonomic-frequency items as having been generated than they are to correctly judge previously generated high-taxonomic-frequency items. In contrast, taxonomic frequency has little or no effect on the number of correct judgments for items that were initially read. This result suggests that more difficult generation operations (presumably those required to generate medium-taxonomic-frequency items) produce more distinctive records of cognitive operations, which are more easily remembered at test.

If these effects of taxonomic frequency on judgments of origin are replicated, then the sensitivity of the judgment of origin data cannot be questioned. In addition, if the read/generate retrieval manipulation produces a replication of the generation enhancement effect in recognition, then the ability of the independent variables to affect recognition cannot be questioned. We can then decisively evaluate the effects of the read/generate retrieval manipulation on judgments of origin.

Method

Subjects. Twenty subjects participated in Experiment 2. The subjects were all students enrolled in Barnard College's Summer Pre-College Program, during the summer between their junior and senior years of high school. Their mean age was 16.5 years. They all volunteered to participate as a class exercise.

Design and Materials. The design of the experiment was a 2 × 2 × 2 completely within-subject design. Read/generate at encoding was crossed with read/generate at retrieval. The third independent variable was the taxonomic frequency (high or medium) of the items.

The stimulus set consisted of four high-frequency (positions 1-6) and four medium-frequency (positions 12-20) words from each of 24 categories in the Battig and Montague (1969) norms. Each subject studied two high-frequency words and two medium-frequency words from each category. The remaining words served as distractors on the recognition test. For each category, one presented word from each frequency level was read and the other was generated. Four different presentation forms were prepared such that each item was read and generated an equal number of times. In addition, those words that were presented to half of the subjects served as distractors for the remaining subjects, and vice versa.

The words were presented in a fixed random order, subject to the constraint that successive members of the same category were separated by a minimum of four intervening items from other categories. Each word was presented along with its category label. In the generate condition, the first three letters of the target word were presented, along with a set of blanks (e.g., fruit-app__; fruit-lem__). No indication as to word length was given.

Two test forms were prepared. On one form, all eight of the items from a randomly selected half of the categories were read. All of the items from the remaining 12 categories were generated at test. This assignment was reversed on the other test form.

Each word was presented along with its category label on the test form in either read or generate form. Next to each item were the words "NEW," "OLD," "READ," and "GENERATE." The items were tested in a fixed random order, subject to the constraint that at least four items intervened between successive members of the same category.

Procedure. All of the subjects were tested at once, in a large room. The subjects were instructed to study each item along with its category label and were told that their memory for these words would be tested. The nature of the test was not specified. For gener-

ate items, they were instructed to use the category labels to generate an instance of the category that began with the specified three letters. They were required to write in the remaining letters for each item that they generated. The subjects worked with a cardboard mask that allowed only one item to show at a time; they moved the mask at 5-sec intervals, which were signaled by brief tones. The subjects were told to skip any words that they could not generate within the interval indicated by the tones.

After the presentation, the subjects were given a short vocabulary test. They were then dismissed. They returned for the test session 48 h later. They were instructed to read or generate each target word on the test as they came to it and to indicate whether it was new (i.e., not seen on the presentation form) or old (i.e., seen as a target word on the presentation form). For items judged old, they were asked to indicate whether they had read the item or generated it on the presentation form. They were instructed to complete all decisions about one item before moving on to the next and not to backtrack. Each subject used the mask, and worked at his/her own pace. The subjects were told that the breakdown of new and old items was roughly half and half.

Results

Items were scored as incorrectly generated if the subject failed to generate the correct word during presentation and/or test. Not surprisingly, incorrect generations were more frequent for medium-taxonomic-frequency items than for high-taxonomic-frequency items. For high-taxonomic-frequency old items, the mean proportions of incorrect generations were .03 for items read at encoding and generated at test, .07 for items generated at encoding and read at test, and .08 for items generated at both encoding and test. The corresponding values for medium-taxonomic-frequency items were .10, .21, and .22. The main effect of taxonomic frequency was significant [$F(1,19) = 48.64$, $MS_e = 0.008$]. There was also a main effect of item type [$F(2,38) = 13.15$, $MS_e = 0.007$]. This effect reflects the fact that subjects are less likely to fail to correctly generate an item at test if it was previously read at encoding than they are to fail to correctly generate an item if it was previously generated at encoding. Incorrect generations for new items that had to be generated at test were also more frequent for medium-taxonomic-frequency items (.21) than for high-taxonomic-frequency items (.07) [$F(1,19) = 35.49$, $MS_e = 0.006$]. It is clear that the likelihood of failing to correctly generate an item on its first encounter did not vary as a function of whether it was first encountered at study or test.

Memory for occurrence. Only the data for correctly generated items were analyzed, using the same procedures used in Experiment 1.

The false-alarm rate was higher for high-taxonomic-frequency items (.34) than for medium-taxonomic-frequency items (.22) [$F(1,19) = 39.58$, $MS_e = 0.007$]. The false-alarm rate for items read at retrieval (.29) was not significantly different from the false-alarm rate for items generated at retrieval (.27) ($F < 1$). The interaction between these two factors was not significant [$F(1,19) = 2.22$].

Because of the false-alarm-rate differences, d' values were computed for each subject in each of the eight conditions. These values are shown in Table 3, along with the uncorrected hit and false-alarm rates. The primary prediction concerns the generation specificity effect. Are items that are initially generated recognized better if they are generated at test than if they are read at test? The answer is affirmative. In addition to a significant main effect of generation at encoding [$F(1,19) = 180.39$, $MS_e = 0.223$] and a nonsignificant main effect of generation at retrieval [$F(1,19) = 4.09$, $MS_e = 0.533$], the read/generate at encoding × read/generate at retrieval interaction was significant [$F(1,19) = 20.95$, $MS_e = 0.181$]. Generating at retrieval, in comparison with reading at retrieval, resulted in a recognition increase of .54 d' units, for those items that were initially generated at encoding. A separate, a priori contrast confirmed that this increase was significant [$F(1,19) = 12.35$, $MS_e = 0.259$]. In contrast, recognition performance for items that were initially read was unaffected by whether they were read or generated at retrieval ($F < 1$).

Table 3
Mean Recognition Hit Rates, False-Alarm Rates, and d' Values for Experiment 2

	Encoding Condition				
	Read		Generate		False-Alarm
Retrieval	Hit Rate	d'	Hit Rate	d'	Rate
	High Taxonomic Frequency				
Read	.62	0.68	.85	1.37	.36
Generate	.57	0.69	.95	1.94	.32
	Medium Taxonomic Frequency				
Read	.62	1.10	.85	1.80	.22
Generate	.56	0.93	.97	2.31	.23

The main effect of taxonomic frequency was also significant [$F(1,19) = 16.32$, $MS_e = 0.326$]. Recognition was better for medium-taxonomic-frequency items ($d' = 1.54$) than for high-taxonomic-frequency items ($d' = 1.17$). This difference was entirely due to the difference in the false-alarm rates for these two types of items; the uncorrected hit rates did not vary as a function of taxonomic frequency. None of the interactions between taxonomic frequency and any of the other factors approached significance (all Fs < 1.06).

Memory for origin. The discrimination (d') and bias (C) scores for the judgment of origin data are shown in Table 4, along with the proportions of correct judgments of origin for each condition. The analyses for the correct judgments were performed on arcsine transformations of the original values. The means of these values were transformed back to proportions, using a sine transformation. These are the values shown in Table 4.

Separate analyses were performed for items that were originally read and for items that were originally generated. First, consider the taxonomic frequency manipulation. As predicted, there was a significant effect of taxonomic frequency for the items that were initially generated [$F(1,19) = 9.06$, $MS_e = 0.083$]. The proportion of correct generate judgments was higher for initially generated medium-taxonomic-frequency items (.92) than for initially generated high-taxonomic-frequency items (.82).

Table 4
Mean Performance Levels for Judgments of Origin in Experiment 2

Retrieval	p("R"/R)*	p("G"/G)†	d'	C	p("R"/N)‡
High Taxonomic Frequency					
Read	.85	.76	1.34	−.08	.80
Generate	.60	.87	1.08	.40	.62
Medium Taxonomic Frequency					
Read	.86	.86	1.67	.01	.73
Generate	.61	.96	1.33	.52	.55

*Probability of responding "read" to an item that was read at encoding, given that it was recognized. †Probability of responding "generate" to an item that was generated at encoding, given that it was recognized. ‡Probability of responding "read" to a new item, given that it was falsely recognized.

In contrast, the proportion of correct read judgments did not differ for high- and medium-taxonomic frequency items (.73 and .75, respectively; $F < 1$).

The effects of generating at retrieval were similar to those found in Experiment 1. Generating at retrieval resulted in a significant (.11) increase in the proportion of correct generate judgments [$F(1,19) = 13.53$, $MS_e = 0.070$]. But it also resulted in a significant (.25) *decrease* in the proportion of correct read judgments [$F(1,19) = 15.34$, $MS_e = 0.183$]. That is, the subjects were simply more likely to respond "generate" to all items when they generated at retrieval.

These conclusions are also reflected in the signal detection analysis. Discrimination of origin was better when the subjects read at retrieval ($d' = 1.51$) than when they generated at retrieval ($d' = 1.21$) [$F(1,19) = 5.32$, $MS_e = 0.337$]. Origin discrimination was also better for medium-taxonomic-frequency items ($d' = 1.50$) than for high-taxonomic-frequency items ($d' = 1.21$) [$F(1,19) = 5.36$, $MS_e = 0.304$]. The interaction between these two factors was not significant ($F < 1$).

The subjects' response bias was also significantly affected by whether they read or generated at retrieval [$F(1,19) = 24.52$, $MS_e = 0.198$]. As described previously, generating at retrieval resulted in a shift in the subjects' response bias towards responding "generate." Response bias was not affected by taxonomic frequency. Both the main effect of this factor and its interaction with the retrieval factor were not significant (both Fs < 1.87).

Another measure of response bias was obtained from the judgments of origin for those new items called old (the false alarms). The mean proportion of read judgments for new items was greater in the read retrieval condition (.77) than in the generate retrieval condition (.59), but this difference was not significant [$F(1,19) = 2.85$]. The mean proportion of read judgments was also higher for high-taxonomic-frequency items (.72) than for medium-taxonomic-frequency items (.64), but this difference was also not significant [$F(1,19) = 2.30$].

Discussion

The recognition results provided another replication of the generation enhancement effect. Generating at test resulted in an increase in recognition performance for items that were originally generated. This effect was found for both high- and medium-taxonomic-frequency instances.

The predicted effects of taxonomic frequency were observed in the judgment of origin data. Taxonomic frequency affected the decisions for items initially generated, but not the decisions for items initially read. Therefore, the effects of the read/generate retrieval manipulation can be evaluated with confidence.

The basic pattern of effects observed in Experiment 1 was replicated. Generating at retrieval resulted in an increase in generate judgments both for items that were initially generated and for items that were initially read. A similar trend was also observed for new items. Overall, origin discrimination accuracy for old recognized items was reduced when the subjects generated at retrieval.

GENERAL DISCUSSION

This research was based on the premise that information about cognitive operations is coded as part of the memory trace. Given a multicomponent view of the memory trace (e.g., Bower, 1967), it is quite possible to access the memory trace and recognize the item, but to be unable to access all of its component attributes. It was hypothesized that generating at retrieval would allow subjects to simply recognize whether those operations had been performed before, and thus provide more reliable information about whether or not information about cognitive operations had been stored as part of the memory trace. Thus, it was predicted that generating at retrieval would increase the accuracy of judgments of origin. This was not observed. The subjects were more likely to claim that both previously read and previously generated items were initially generated. Overall, judgments of origin were less accurate when subjects generated at retrieval than when they read at retrieval.

Typically, recall of information about cognitive operations provides a quite reliable cue for indicating that the item was generated (Finke, Johnson, & Shyi, 1988; Johnson et al., 1980; Johnson et al., 1981; Rabinowitz, 1989). But when subjects generate at retrieval, it is hypothesized that they find it quite difficult to distinguish between the operations they just completed and the retrieved record of encoded operations. When subjects are faced with this difficult recognition of operations decision, they are quite likely to judge the just performed operations as having been performed previously. This may be because many of the operations performed on different items are quite similar, and recognition of a particular set of operations may need to be based on rather subtle cues, such as the speed with which the operations can be performed. In essence, generating at retrieval results in confusing the just-performed operations for those in the memory trace. This high "false-alarm rate" for new cognitive operations reduces the validity of information about cognitive operations for the origin judgments, and it results in a decrease

in discrimination accuracy for these judgments. That is, mistaken attributions about the source of the operations result in additional errors about the origin of the memory.

This hypothesis also accounts for the dramatic changes in response bias that resulted from generating at retrieval. Previous work in this paradigm, in which the target items were always read at retrieval, has consistently found a read bias in young adult subjects' judgments (Johnson et al., 1980; Johnson et al., 1981; Rabinowitz, 1989). Johnson and her colleagues, and Anderson (1984), explained this bias by positing that subjects selectively search for information regarding cognitive operations for judging an item's origin. Failure to retrieve such information results in a read decision by default. In the present experiments, a read bias was also generally observed when the subjects read at retrieval. However, this bias was eliminated or reversed when the subjects generated at retrieval. Generating at retrieval increased the probability of responding "generate" both to items that were initially generated and to those that were read. It also resulted in an increase in the proportion of generate responses to items that had not been initially studied. If subjects primarily rely on information about cognitive operations to determine that a word was generated, and they confuse the operations just performed during the test for information retrieved from the memory trace, then the large bias shift would be expected.

By focusing on the decision processes involved in evaluating information about cognitive operations, the Johnson and Raye (1981) reality monitoring framework appears to be able to give a good account of the effects of generating at retrieval on origin judgments. We also need to account for the rather different effects observed from this manipulation on recognition. Repeating the same generation operations at retrieval as were performed at encoding facilitates recognition. In contrast, there is no effect of generation at retrieval on recognition performance if the items were originally read, or if the generation operations are different from those originally performed.

First, it should be noted that a proceduralist view of memory (Kolers & Roediger, 1984) is quite compatible with the recognition results, but not with the judgment of origin results. Kolers and his colleagues have argued against the notion of a static memory trace that includes lists of encoded features (such as the source, or origin, of the item). Rather, they have argued for a procedural representation. The operations engaged during encoding are themselves the record of the event. Recognition (memory for occurrence) is held to be a function of the degree to which the operations required at test match those initially performed. But if what is actually stored about a stimulus event are the operations performed on the event, then memory for occurrence should inherently bring with it the source or the procedures used to encode the item. By itself then, this view would not predict confusions about the source of information about cognitive operations.

Alternatively, the headed records memory model (Morton, Hammersley, & Bekerian, 1985) can account for the different effects of generation at retrieval on recognition and judgments of origin. According to this model, events are stored as individual, unconnected records, to each of which is attached a heading. Access to a memory record is achieved through a match between a search description and the information in the heading. There are two critical assumptions of this model. First, the information contained in the heading is different from the information in the record itself. Headings contain relatively unprocessed information and literal features of the stimulus and the environment. Records contain more "processed" information. Second, accessing a record via its heading makes accessible all the information in the record. However, the information in the heading itself is inaccessible and not subject to recall.

According to this model then, we would posit that the cognitive operations performed when generating an item are stored in a relatively unprocessed state as part of the heading of the memory record for that item. When these generation operations are repeated at test, they form a part of the search description and thus increase the chances of matching the memory heading, resulting in recognition of the item. However, only the information in the record can be retrieved, and it is therefore that information that must be used to make the origin decision. Of course, some information about the cognitive operations performed during encoding must be part of the memory record, or we would be unable to account for the previously reported findings that both the repetition and the complexity of the generation operations at encoding affect judgments of origin. It is likely that this information in the record is stored in a form somewhat different from the information in the heading. In any event, the most important aspect of the headed records model, in terms of our data, is that, once the record is accessed, all of the information in it is accessible, and this is independent of the means by which it was accessed (the particular information in the description that matched the information in the heading).

Once the information in the record is accessed, it must be evaluated. It is here that the judgment processes for reality monitoring, described previously, come into play. One direction for future research is to determine why the "recognition of operations test" induced by our generation at retrieval task was so difficult, and whether other types of cognitive operations may be more discriminable. If such operations can be found, generating at test may yet improve discrimination of origin.

REFERENCES

ANDERSON, R. E. (1984). Did I do it or did I only imagine doing it. *Journal of Experimental Psychology: General*, **113**, 594-614.

BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, **80**, 3, pt. 2.

BOWER, G. (1967). A multi-component theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 1, pp. 229-325). New York: Academic Press.

CERMAK, G., SCHNORR, J., & BUSCHKE, H. (1970). Word recognition as a function of spelling direction during study and test. *Psychonomic Science*, **21**, 127-128.

DONALDSON, W., & BASS, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning & Verbal Behavior*, **19**, 26-35.

FINKE, R. A., JOHNSON, M. K., & SHYI, G. C.-W. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, **16**, 133-137.

GLISKY, E. L., & RABINOWITZ, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 193-205.

GREENWALD, A. G., & JOHNSON, M. M. S. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition*, **17**, 673-681.

HIRSHMAN, E., & BJORK, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 484-494.

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

JOHNSON, M. K., RAYE, C. L., & DURSO, F. T. (1980). Reality monitoring: Second thoughts and perceptions. *Bulletin of the Psychonomic Society*, **15**, 402-404.

JOHNSON, M. K., RAYE, C. L., FOLEY, H. J., & FOLEY, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, **94**, 37-64.

KOLERS, P. A. (1975). Specificity of operations in sentence recognition. *Cognitive Psychology*, **7**, 289-306.

KOLERS, P. A. (1979). A pattern-analyzing basis of recognition. In L. S. Cermak and F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.

KOLERS, P. A., & ROEDIGER, H. L. III. (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior*, **23**, 425-449.

MANDLER, G., RABINOWITZ, J. C., & SIMON, R. A. (1981). Coordinate organization: The holistic representation of word pairs. *American Journal of Psychology*, **94**, 209-222.

MORTON, J., HAMMERSLEY, R. H., & BEKERIAN, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, **20**, 1-23.

NAIRNE, J. S., PUSEN, C., & WIDNER, R. L., JR. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition*, **13**, 183-191.

NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.

RABINOWITZ, J. C. (1989). Judgments of origin and generation effects: Comparisons between young and elderly adults. *Psychology & Aging*, **4**, 259-268.

RABINOWITZ, J. C., & CRAIK, F. I. M. (1986). Specific enhancement effects associated with word generation. *Journal of Memory & Language*, **25**, 226-237.

RUSSO, J. E., & WISHER, R. A. (1976). Reprocessing as a recognition cue. *Memory & Cognition*, **4**, 683-689.

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

SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.

THOMSON, D. M. (1972). Context effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior*, **11**, 497-511.

TULVING, E., & THOMSON, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, **87**, 116-124.

VOSS, J. F., VESONDER, G. T., POST, T. A., & NEY, L. G. (1987). Was the item recalled and if so by whom? *Journal of Memory & Language*, **26**, 466-479.

NOTE

1. A number of other methods of analysis led to the same conclusions. The effects of reading or generating at retrieval on discrimination accuracy were observed with both the Goodman-Kruskal gamma correlation (see Nelson, 1984; Voss, Vesonder, Post, & Ney, 1987) and an analysis of a posteriori probabilities (see Anderson, 1984). The Goodman-Kruskal gamma correlation was used to assess the concordance between an item's status (read or generated at encoding) and the subject's response to it. It was significantly higher in the read retrieval condition (.85) than in the generate retrieval condition (.79). In addition, the a posteriori probability of being correct given a generate response was significantly lower in the generate retrieval condition (.83) than in the read retrieval condition (.88). The a posteriori probability of being correct given a read response did not differ between the read and generate retrieval conditions (.70 and .71, respectively). The bias effects observed with *C* were also observed in an analysis of simple probabilities. The overall probability of responding "read" to all recognized old items was .63 in the read retrieval condition; it was significantly lower (.54) in the generate retrieval condition.

(Manuscript received March 23, 1988;
revision accepted for publication June 5, 1989.)