Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 5, 858–871

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.5.858

Covert Operations: Orthographic Recoding as a Basis for Repetition Priming in Word Identification

Michael E. J. Masson
University of Victoria

Colin M. MacLeod
University of Toronto at Scarborough

When a word is generated from a semantic cue, coincident orthographic visualization of that word may cause priming on a subsequent perceptual identification test. A task was introduced that required subjects to visualize the orthographic pattern of auditorily presented words. When used at study, this task produced a pattern of priming similar to that produced by a generate study task. When used at test, equal priming on the orthographic task was produced by read and generate study tasks but not by a generate study task that failed to invite orthographic visualization. Priming on perceptually based word identification tests that results from a generate study episode may be largely due to orthographic recoding of the target rather than to conceptual processing.

The beneficial effects of prior encoding of a stimulus on subsequent encoding of that stimulus are generally governed by the principles that characterize the transfer-appropriate processing framework (Morris, Bransford, & Franks, 1977; Roediger, 1990; Roediger, Weldon, & Challis, 1989). In particular, repetition-priming benefits are most powerful when the processing operations applied during the initial encoding of a stimulus are most similar to those applied when the stimulus is presented on a later occasion. So, for example, completion of visually presented word fragments with target words is more likely if those targets were previously experienced as printed words rather than as auditorily presented words or as drawings of objects corresponding to those words (Rajaram & Roediger, 1993).

Another important example of the principles of transfer-appropriate processing is Jacoby's (1983) demonstration that although prior visual exposure to words led to enhanced accuracy on a subsequent masked word identification task, prior generation of those items from their antonyms, in which the target words were not actually seen, produced little or no enhancement. Toth, Reingold, and Jacoby (1994), Weldon (1991), and Jacoby, Toth, and Yonelinas (1993) reported similar results using other forms of target generation at study, such as definitional phrases or anagrams, followed by a word-stem or word-fragment completion test or a masked word identification test. Despite the ineffectual contributions of generate encoding tasks to apparently perceptually based word identification tests, these encoding tasks consistently produce reliably better performance on tests that are strongly supported by conceptually driven processing, such as recognition and generation of category members (Blaxton, 1989; Jacoby, 1983; Jacoby et al., 1993; Toth et al., 1994; Weldon, 1991).

The Problem of Priming Due to Generation

A persistent exception to this coherent pattern of results has been the finding that under a rather wide range of conditions, generation of target words from semantic cues is sufficient to produce as much repetition priming on a subsequent masked word identification test as is produced by an encoding task that provides subjects with direct visual experience of the targets (MacLeod & Masson, 1997; Masson & MacLeod, 1992, 1997). For example, Masson and MacLeod (1992) reported that generating targets from such cues as definitions or synonyms and associates, without actually seeing the targets, led to as much repetition priming as did reading targets aloud. We proposed that the masked word identification test involves a component of conceptual processing—recruitment of word meaning as part of the word-identification process—that benefits from prior conceptual encoding episodes. Given the typical classification of the masked word identification test as perceptually based, our pattern of results is anomalous in the context of the standard transfer-appropriate processing framework.

An alternative explanation for the repetition-priming effects following generate encoding tasks is that they arise from conscious recollection of previously encoded items. Jacoby et al. (1993) and Toth et al. (1994) argued that priming on indirect tests of memory is multiply determined—with both conscious and unconscious influences affecting performance. To estimate these two kinds of influence separately, they applied the process-dissociation procedure to analyze priming on a word-stem completion test. Their estimates of conscious and unconscious influences, based on the assumption that these sources have independent influences, revealed that nonperceptual encoding tasks such as anagram solution or generation from a semantic cue induced no unconscious influ-

The research reported here was supported by Natural Sciences and Engineering Research Council of Canada Research Grants A7910 and A7459 to Michael E. J. Masson and Colin M. MacLeod, respectively. We extend thanks to Sahira Iqbal, who conducted Experiments 1 and 2 for her bachelor's thesis at the University of Victoria, and to Allison Barnes and Lucas Riedl for their assistance in carrying out these experiments.

Correspondence concerning this article should be addressed to Michael E. J. Masson, Department of Psychology, University of Victoria, P.O. Box 3050, STN CSC, Victoria, British Columbia V8W 3P5, Canada, or to Colin M. MacLeod, Division of Life Sciences, University of Toronto at Scarborough, Scarborough, Ontario M1C 1A4, Canada. E-mail: mmasson@uvic.ca or macleod@utsc.utoronto.ca

ence on word-stem completion. Priming by these tasks on perceptually based indirect tests of memory would therefore be due to conscious recollection.

In contrast to these conclusions, Bodner, Masson, and Caldwell (2000) showed for the word-stem completion test that application of process-dissociation equations on the basis of the assumption of independent conscious and unconscious influences may lead to underestimation of the unconscious influence of generate encoding tasks. Underestimation arises because the high level of conscious recollection engendered by generate encoding tasks can reduce the use of target completions under exclusion instructions without creating an offsetting increase in the use of target completions under inclusion instructions. Bodner et al. showed that having subjects read and provide an associate to a target word at study led to much lower estimates of unconscious influences on a later word-stem completion test than did having subjects simply read a target word at study. The added semantic component of the read–associate study task reduced the use of target completions under exclusion instructions while producing little change in inclusion performance. Thus, although read and read–associate encoding tasks afford identical perceptual experiences, they lead to very different estimates of unconscious influences because of the semantic component of the read–associate task. A similar underestimation could be responsible for the finding that generate and anagram encoding tasks fail to produce unconscious influences on word-stem completion when those influences are estimated by process-dissociation equations (Bodner et al., 2000; Jacoby et al., 1993; Toth et al., 1994).

Attribution of generation-based priming on purportedly perceptual tests to conscious recollection is also inconsistent with the finding that a generate encoding task can produce significant priming on a speeded word-reading test (MacLeod & Masson, 2000). On such a speeded test, paying heed to conscious influences of memory tends to work against making a rapid response; therefore, to optimize performance it actually would be better to ignore these influences rather than to act on them (Horton, Wilson, & Evans, 2001; MacLeod & Masson, 2000).

Other evidence also argues against the proposal that generation-based priming on perceptual indirect tests is due to conscious recollection. For example, McDermott (1997) and McKone and Murphy (2000) demonstrated that studying a list of conceptually related words (e.g., *tired*, *bed*, *dream*) that does not include the thematically central word (i.e., *sleep*) nevertheless led to priming of that central word on a word-stem completion test. Moreover, McKone and Murphy found this priming effect to be modality specific, inconsistent with it being the result of conscious recollection.

Finally, Masson (in press) used a two-alternative forced-choice masked word identification task in which one of the alternatives had been presented at study to show that read and generate encoding tasks can produce similar amounts of bias toward the studied alternative. That bias effect was eliminated, however, when the members of a pair of probe words were orthographically different (see also Ratcliff & McKoon, 1997), indicating that the effect was not due to conscious recollection. If it had been, then the bias should have emerged regardless of whether probes were orthographically similar or different.

Could Generation Involve Orthographic Recoding?

There is a third reason that generate encoding might produce priming on a perceptual test. During the encoding task, subjects may engage in a form of covert orthographic processing of target words, in which they visualize a target's printed appearance. There is evidence that auditory encoding tasks that require analysis of a word's component phonemic structure induce subjects to engage in orthographic recoding of this kind (Neaderhiser & Church, 2000; Seidenberg & Tanenhaus, 1979). For example, Seidenberg and Tanenhaus showed that with auditorily presented words, latency to decide that a target word rhymed with a cue word increased if the two items were orthographically different rather than similar (e.g., *rye–tie* vs. *pie–tie*). Orthographic recoding may also occur when subjects generate a target word from a semantic cue, perhaps as part of the process of verifying that a candidate word fits the cue. Moreover, the quite common inclusion of the first letter (MacLeod & Masson, 1997; Masson & MacLeod, 1992; Toth et al., 1994; Weldon, 1991) or of several letters (Hirshman, Passannante, & Arndt, 2001; Toth & Hunt, 1990) from the target as part of the generation cue might encourage orthographic recoding by inviting analysis of a target's phonemic structure (i.e., its fit to the available letter or letters in the cue).

There are both theoretical and empirical grounds for arguing that internally generated orthographic representations can influence priming on perceptual tests. First, there is evidence that similar neural systems are involved in both visual perception and visual imagery (Farah, 1988; Farah, Peronnet, Gonon, & Girard, 1988; Kosslyn et al., 1993). Second, it has been found that generating a mental image of a word or spelling an auditorily presented word leads to priming on word identification tests and that forming a mental image of a named object primes later object identification (e.g., Jacoby & Witherspoon, 1982; McDermott & Roediger, 1994; see also Roediger & Blaxton, 1987; Roediger, Weldon, Stadler, & Riegler, 1992; Schacter & Graf, 1989). Moreover, these priming effects are modality specific (e.g., forming a mental image of a word leads to more priming on a word identification test than does seeing the corresponding object), indicating that they are the product of perceptually driven processes.

The fact that deliberate attempts to visualize the appearance of a stimulus can lead to later priming supports the suggestion that orthographic recoding associated with a generate encoding task might be responsible for priming rendered by that task. This explanation for priming induced by generate encoding tasks is appealing from a transfer-appropriate processing perspective because it fits with the assumption that tests such as masked word identification and word-stem completion are perceptually, not conceptually, driven.

On this account, priming is due either to actually seeing a word or to imagining its orthographic form. To test this account, we developed a new indirect test of memory—the letter-height task—that requires subjects to image a word's orthographic pattern in response to an auditory presentation of that word. More specifically, the task requires subjects to imagine a word's lowercase orthography and to use that representation to determine the number of ascending (e.g., *f*, *t*) and descending (e.g., *g*, *p*) letters in the word. The target word is thereby recoded into an orthographic form despite not having been actually seen.

In the eight experiments reported here, we used the letter-height task as a means of assessing orthographic processing, both as an encoding task and as an indirect test of memory. An advantage of the letter-height task as an encoding task over instructions to imagine a word's orthography as an encoding task (e.g., McDermott & Roediger, 1994; Roediger et al., 1992) is that performance accuracy can be monitored to ensure that subjects are actually engaging a word's orthographic pattern. Using the letter-height task at encoding permitted us to assess the extent of priming on subsequent indirect tests of memory that could be expected from an encoding task involving visualization of orthography. In turn, the pattern of priming effects across various indirect tests of memory produced when the letter-height task is used at encoding can be compared with the pattern of priming caused by a generate encoding task. If these two patterns are very similar, then the orthographic-recoding hypothesis will gain support.

We also used the letter-height task as an indirect test of memory to show that when serving as a memory test, this task is specifically sensitive to prior orthographic encoding experience. That specificity sets the stage for a convincing test of the proposal that generate encoding tasks lead to priming on visual word identification tasks in large part because of covert orthographic processing of generated targets.

Experiment 1

Before examining (a) the patterns of priming produced by letter-height and generate encoding tasks and (b) the priming effects obtained with the letter-height task as an indirect test of memory, we examined the possibility that the modality in which generate cues are presented might influence priming on a subsequent indirect test of memory. There are two reasons to be concerned about such a possibility. First, one of our primary goals is to compare priming effects produced by the letter-height task and a generate task; these tasks typically involve different presentation modalities. If these tasks lead to different amounts of priming, it is important to know whether modality of presentation of the encoding stimulus is responsible. Establishing that auditory and visual generate cues yield similar amounts of priming would alleviate this concern. Second, simply from the standpoint of the match between study and test, processing induced by auditory generation cues might not lead to as much priming on a subsequent visual identification test as would processing induced by visual generation cues. Therefore, in Experiment 1, we compared priming on a masked word identification test following visual versus auditory generate cues.

Method

Subjects. Forty-eight students from an introductory psychology course at the University of Victoria, Victoria, British Columbia, took part in the experiment for extra credit in their course. Half of the subjects were randomly assigned to each of the two study modality conditions.

Materials. A set of 90 critical and six practice items, each consisting of a phrase cue and corresponding target word (e.g., "The alcohol produced when grapes ferment"–"wine"), was selected from the set of materials used by Masson and MacLeod (1992). The normative frequency of the target words ranged from 1 to 492 per million, with a median frequency of 27 (Kučera & Francis, 1967). Word length ranged from four to six letters. The 90 critical items were divided into three lists of 30. Assignment of these lists to the three study conditions—read, generate, and new—was counterbalanced across subjects so that each item appeared equally often in each study condition. Each list of 30 cues, including the first letter of the target word (e.g., "The visual imagery that occurs during sleep"–"d") was recorded by a female speaker on an audio tape for use in the auditory generate condition. An additional set of 45 words ranging from four to six letters was selected for use in the threshold-setting phase of the experiment.

Procedure. Subjects were tested individually in a quiet room using an Apple Macintosh II (Cupertino, CA) desktop computer equipped with two monochrome monitors and a cassette tape recorder. One monitor was used to present instructions and stimuli to the subject, with stimuli appearing as black letters against a white background. The other monitor displayed information to the experimenter about the target stimulus on each trial, allowing determination of the accuracy of the subject's responses.

In the study phase, subjects were presented with two blocks of trials, each involving a different encoding task. Order of task presentation was counterbalanced across subjects. For the read task, three practice words and then 30 critical target words were presented. On each trial, the word appeared in lowercase letters at the center of the monitor and remained in view until the subject read it aloud. The experimenter then pressed a key on the computer keyboard to initiate the next trial. For the generate task, there were two versions. For subjects in the visual group, each generation cue was presented at the center of the computer monitor and remained in view until the subject responded. Each cue consisted of a descriptive phrase and the first letter of the target word followed by a question mark (e.g., "Something one uses to sit at a desk"–"c?"). For subjects in the auditory group, the cues were identical to those presented in the visual version of the task, but they were presented auditorily using a cassette tape recorder. In both the visual and auditory versions of the generate task, the experimenter pressed a key to classify the correctness of the response and orally provided the correct response if the subject failed to produce it.

Each subject was then presented with a series of threshold-setting trials for a masked word identification test. On each trial, the word *READY* appeared at the center of the computer monitor as a warning signal and remained in view until the subject pressed a key mounted on a response box. The warning signal was then erased, and a word was presented for 30 ms. The target word was immediately followed by a mask consisting of a row of six @ symbols. The subject then attempted to report the target, the experimenter recorded the response, and the next trial began automatically. During this phase, the contrast (blackness) of the displayed targets was adjusted on the basis of the subject's rate of correct responding. Adjustments were made using a titration estimation procedure (Pentland, 1980) to achieve a proportion correct of .5. The contrast level reached during this procedure was used for all critical test trials.

The test phase used the same procedure as the titration phase. Ninety critical target words, 30 from each study condition and 30 nonstudied, were presented in random order. Thus, of the 90 critical items, .67 were studied. But there was no interruption between the 45 titration items and the 90 critical items, so the overall proportion of studied items presented for masked word identification was .44 (60 of 135). In this and all other experiments reported here, subjects were not informed that items from the study phase would appear on this test, although they likely noticed some of these repetitions.

Results and Discussion

The mean proportions of words correctly generated during the study phase in the visual and auditory groups were .94 and .90, respectively. Data from the test phase were analyzed both by conditionalizing on correct generation during the study phase and without conditionalizing. The pattern of results was the same in both cases, therefore only the results from the unconditionalized data are reported. This approach was followed for all of the experiments reported here. The probability of a Type I error was set at .05 for all analyses reported in this article.

Mean proportions of correctly identified target words in the masked word identification test are shown in Figure 1. These data were submitted to an analysis of variance (ANOVA) with modality of the generation cues (visual and auditory) as a between-subjects factor and encoding task (read, generate, and nonstudied) as a within-subject factor. The effect of encoding task was significant, $F(2, 92) = 27.29$, $MSE = .010$. The effect of modality of generation cues approached significance, $F(1, 46) = 3.43$, $MSE = .112$, $p = .07$. The interaction of these two factors was not significant ($F < 1$). Figure 1 indicates that the read and the generate encoding tasks produced similar priming effects relative to the nonstudied condition. This pattern held for both modality groups. Pairwise comparisons collapsing across modality groups and using the interaction error term from the ANOVA showed that although read and generate encoding conditions did not differ reliably ($d = .03$), $F(1, 92) = 2.28$, both produced priming relative to the nonstudied condition, $F(1, 92) = 49.25$; and $F(1, 92) = 30.35$, for read ($d = .14$) and generate ($d = .11$), respectively.

It is not clear why there was a trend for greater accuracy when subjects were given the generation cues in an auditory format; the two groups had similar contrast levels for the masked word identification test, and subjects were randomly assigned to modality conditions. We suspect that the result is due to happenstance and, in any case, modality did not interact with the magnitude of observed priming effects.

These results replicate our earlier findings of comparable priming effects for the read and generate encoding tasks using masked word identification as the test (e.g., MacLeod & Masson, 1997; Masson & MacLeod, 1992). Moreover, Experiment 1 also demonstrates the new fact that the priming observed in that test situation is not dependent on the modality in which generation cues are presented. It is interesting to note that this outcome is not consistent with the proposal (McKenna & Warrington, 1993) that conceptual knowledge is modality specific, although we certainly do not consider Experiment 1 a particularly strong test of that hypothesis.

Having established that auditory and visual generation cues are equally potent, one can assume that regardless of whether the encoding operations that cause generation-based priming derive

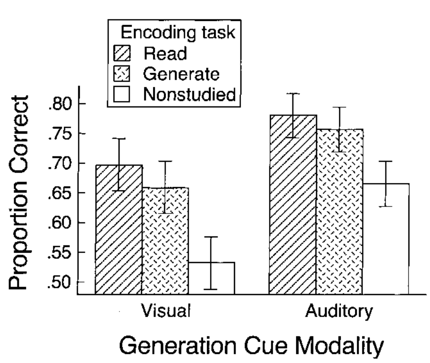

Figure 1. Experiment 1: Mean proportion correct in the masked word identification test as a function of generation cue modality and encoding task. Error bars show the 95% within-subject confidence intervals for each modality group (Loftus & Masson, 1994).

from covert generation of a target's orthography or are conceptual in nature, they appear to be independent of the modality in which the generation cue is presented. With this point established, we move to a series of experiments that tests the orthographic-recoding hypothesis using the novel letter-height task expressly designed for this purpose.

Experiment 2

The encoding operations that produce priming for generated targets are not modality dependent. This helps to justify the switch to the auditory presentation of stimulus materials required for the letter-height task, the new task designed to induce orthographic processing. In Experiment 2, we contrasted visual presentation requiring reading with auditory presentation requiring one of two responses: simply repeating aloud or performing the letter-height task. We reasoned that if the orthographic-recoding hypothesis is correct, then to the extent that repeating aloud engages only acoustic processing, there should be little or no priming in that condition. In contrast, on the basis of its orthographic requirement, there should be priming in the letter-height encoding condition, and it should be of similar magnitude to the priming due to reading.

Method

Subjects. Forty-eight subjects were drawn from the same pool as in Experiment 1. Half of the subjects were randomly assigned to each of the two study conditions.

Materials and procedure. The items from Experiment 1 were used, except that two critical words that were homophones were replaced with nonhomophonic words. Two of the three sets of 30 critical items were presented in the study phase, with the third set of 30 critical items assigned to the nonstudied condition. In the study phase, one set of items was presented in the read task, and the other set was presented in the repeat-only task or the letter-height task, depending on the group to which the subject had been assigned. Assignment of sets of 30 critical items to these study conditions was counterbalanced across subjects.

For subjects in the repeat-only group, the auditory task involved two parts: listening to the target word and repeating it aloud. For subjects in the letter-height group, the task involved four parts: (a) listening to the target word, (b) repeating the word aloud, (c) forming a mental image of the word in lowercase letters, and (d) reporting the number of letters in the word that had ascending (*b, d, f, h, k, l, t*) or descending (*g, j, p, q, y*) features. Critical words were recorded in a female voice on the computer and presented over the computer's built-in speaker. In the study phase, the two types of study trials (read and repeat only/letter height) were mixed and presented in random order. We have previously shown that mixing versus blocking has no impact when just two encoding tasks are involved (MacLeod & Masson, 1997). The test phase was conducted using the same procedure as in Experiment 1, including 45 titration trials. Thus, across the entire test phase, the proportion of studied items presented was .44.

Results and Discussion

In the study phase, the mean proportion of words to which subjects in the letter-height group responded incorrectly (either through erroneous repetition or erroneous letter count) was .06. Subjects in the repeat-only group made no errors in repeating the auditory targets.

The mean proportions of correct identification responses in the test phase are shown in Figure 2. An ANOVA with auditory task

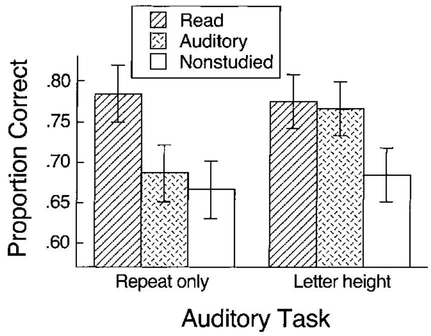

Figure 2. Experiment 2: Mean proportion correct in the masked word identification test as a function of auditory task group and encoding task. Error bars show the 95% within-subject confidence intervals for each auditory task group.

(repeat only, letter height) as a between-subjects factor and encoding task (read, auditory, nonstudied) as a within-subject factor revealed a main effect of encoding task, $F(2, 92) = 18.69$, $MSE = .007$; and an interaction between auditory task and encoding task, $F(2, 92) = 3.54$, $MSE = .007$. The main effect of auditory task was not significant ($F < 1$). Given the reliable interaction, we applied pairwise comparisons separately to the two auditory-task groups. When the auditory task was to repeat the target words, no priming was obtained ($d = .02$; $F < 1$), although reading visually presented targets resulted in substantial priming ($d = .12$), $F(1, 46) = 22.76$, $MSE = .007$. When the letter-height task was applied to auditory targets, however, there was reliable priming ($d = .08$), $F(1, 46) = 12.04$, $MSE = .007$, but no reliable difference between the letter-height and read conditions ($d = .01$; $F < 1$).

In Experiment 1, we showed that robust priming is obtained when an auditory study task also requires target generation. In contrast, in Experiment 2, the lack of priming with the standard auditory study task (listen and repeat) replicates earlier demonstrations using the masked word identification test (Jacoby & Dallas, 1981; Weldon, 1991). We hypothesized, however, that the letter-height task would engage orthographic processing and that the resultant recoding potentially could support priming on the masked word identification test. Indeed, in Experiment 2, the letter-height task led to just as much priming as an actual visual presentation of the target, even though the letter-height task involved auditory presentation of target words. This outcome parallels the results we have found for generation tasks, both in Experiment 1 here and in earlier work (MacLeod & Masson, 1997; Masson & MacLeod, 1992).

Experiment 3

One of our primary goals was to compare priming when encoding is done in the context of the letter-height task with priming when encoding is done in the context of the generation task. Under the orthographic-recoding hypothesis, each of these encoding tasks should covertly produce the orthographic encoding that supports the priming observed when encoding is done by reading. Thus, both letter-height judgment and generation should have produced priming equivalent to that due to reading. But there is the ever-present concern that conscious recollection may contaminate indirect test performance. A straightforward comparison of the influence of these two encoding tasks on a direct test would, therefore, be instructive before turning to a comparison of these tasks in the context of indirect tests.

It is well established that generation produces very good performance on direct tests (e.g., Masson & MacLeod, 1992; Slamecka & Graf, 1978), in which conceptual processing of the studied words is emphasized. The letter-height task, in contrast, would appear not to be conceptual, with its emphasis on structural aspects of the studied words; indeed, its structural emphasis is the primary virtue of the letter-height task for our purposes. Therefore, we expected considerably better performance on a direct test following generation than following letter-height judgment. Experiment 3 used a recognition test to determine whether this was in fact the case.

Method

Subjects. Fifteen students were drawn from the same pool as in the earlier experiments.

Materials. A set of 90 critical target words and generation cues, most drawn from those used in the earlier experiments, was constructed. The targets varied in word frequency from 0 to 492, with a median of 26 (Kučera & Francis, 1967). The cues were of the same kind as those used in the earlier experiments (e.g., "The granular material found on a beach"–"s"). These items were divided into three sets of 30 items for counterbalanced assignment to three study conditions: generate, letter height, and nonstudied. Each set was used equally often in each condition. Eight additional items were selected for use on practice trials in the study phase.

Procedure. Subjects in Experiment 3 and in subsequent experiments were tested using an Apple Macintosh G3 desktop computer. The same general procedure as in the earlier experiments was used. In the study phase, subjects were presented eight practice trials (four generate and four letter-height) followed by 30 critical trials in the generate task and another 30 trials in the letter-height task. Presentation of these two tasks was mixed, with items presented in random order. In the test phase, a recognition test was presented in which 90 test words were shown one at a time at the center of the computer monitor. The proportion of studied items in the test phase was .67. Subjects were instructed to classify each item as having been presented in the study phase, either in the generate task or in the letter-height task, or as new. Responses were made by pressing one of two designated buttons mounted on a response box.

Results and Discussion

The mean proportions of correct responses in the study phase were 1.00 and .90 for the generate and letter-height tasks, respectively. The mean proportion of generate items correctly classified as previously presented was .88 ($SD = .06$), whereas for the letter-height items the mean proportion recognized was .49 ($SD = .20$). This difference was statistically significant, $F(1, 14) = 66.33$, $MSE = .017$. When performance on the letter-height items was conditionalized on correct responding in the study phase, the mean proportion of recognized items increased to .56 ($SD = .24$), but this value was still reliably less than that of the generate condition, $F(1, 14) = 30.49$, $MSE = .025$. The mean proportion of false alarms was .05.

Consistent with substantial literature (e.g., Masson & MacLeod, 1992; Slamecka & Graf, 1978), the generate encoding task again resulted in very good performance on a direct test of recognition. More crucial for our purposes, recognition was markedly better following generation than following letter-height judgment. This is consistent with our view of letter-height judgment as involving very little conceptual encoding relative to generation. Any account of the priming that we report in subsequent experiments that rests on the intrusion of conscious recollection—or that relies on conceptual processing playing a pivotal role—would therefore have to either anticipate greater priming from the generate encoding task than from the letter-height encoding task or assume that priming from these two tasks arises from different sources. The latter possibility suggests that one should observe divergent patterns of priming for generate and letter-height encoding tasks. As we repeatedly show, however, these two tasks produce comparable levels and patterns of priming. Therefore, for the reasons outlined in the introduction, and given the pattern just reported on a direct test, we do not see conscious recollection as a likely explanation of our findings. We now return to our primary series of experiments.

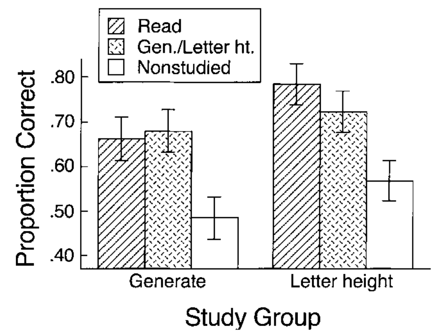

Figure 3. Experiment 4: Mean proportion correct in the masked word identification test as a function of study group and encoding task. Error bars show the 95% within-subject confidence intervals for each study group. Gen. = generate; Letter ht. = letter height.

Experiment 4

Experiment 4 took the opposite tack to Experiment 2, again using masked word identification as the indirect test. Rather than showing a different pattern of priming for auditory presentation with repetition versus the letter-height task, our aim was to show the same pattern of priming for the generate task and the letter-height task. This outcome would be consistent with the hypothesis that these two encoding tasks produce priming on the same basis: orthographic recoding.

Method

Subjects. Forty-eight subjects drawn from the same pool as in the preceding experiments were tested. Half of the subjects were randomly assigned to each of the two study groups. One group was given the generate and read encoding tasks; the other group was given the letter-height and read encoding tasks.

Materials. The practice items and the 90 critical items from Experiment 3 were used. The critical items were divided into three sets of 30 items for counterbalanced assignment to three study conditions: read, generate/letter height, and nonstudied. Each set was used equally often in each condition. An additional set of 45 words was constructed for the titration procedure that was used to set the contrast level for individual subjects in the masked word identification test.

Procedure. The same general procedure as in the earlier experiments was used. In the study phase, the generate group was presented 30 items in the read task and another 30 items in the generation task. Presentation of these two tasks was mixed, with items presented in random order. A similar procedure was followed for the letter-height group, except that these subjects performed the letter-height task rather than the generate task for one set of 30 items. In the test phase, the masked word identification test began with a 45-trial titration procedure, followed by 90 critical trials. Because all items shown on the titration trials were nonstudied, the proportion of studied items shown across the entire test phase was .44.

Results and Discussion

In the study phase, the mean proportion of targets correctly generated was .94; the mean proportion of correct responses to targets in the letter-height task was .93. Mean proportions correct in the masked word identification test are shown in Figure 3. An ANOVA of test performance with study group (generate, letter height) as a between-subjects factor and encoding condition (read, generate/letter height, nonstudied) as a within-subject factor revealed a main effect of encoding condition, $F(2, 92) = 42.67$, $MSE = .013$. Neither the main effect of study group, $F(1, 46) = 1.91$, $MSE = .129$, nor the interaction, $F(1, 92) = 1.43$, $MSE = .013$, was significant. The pattern of means in Figure 3 indicates that both the generate task and the letter-height task produced priming and that the amount of this priming for each of these tasks was about the same as that yielded by the read task. Pairwise comparisons collapsing across the two study groups and using the interaction error term from the ANOVA supported these conclusions, showing that the generate and letter-height conditions combined were significantly different from the nonstudied condition ($d = .17$), $F(1, 92) = 56.10$, but did not differ from the read condition ($d = .02$; $F < 1$). These findings are consistent with the hypothesis that the priming produced by the generate and letter-height tasks is based on a common process, namely, covert orthographic processing of unseen targets.

Although the ANOVA results indicate that the letter-height and generate encoding tasks each produced as much priming as the read task, it is apparent from Figure 3 that for the letter-height study group, accuracy was somewhat lower in the letter-height condition than in the read condition ($d = .06$). By contrast, for the generate study group, accuracy in the read and generate conditions was very similar ($d = .02$). An ANOVA with study group as a factor and two levels of encoding task (read vs. generate/letter height) as the other factor revealed that the interaction of these two factors approached significance, $F(1, 46) = 3.18$, $MSE = .012$, $p < .10$. Nevertheless, the amount of priming produced by the letter-height and generate tasks was very similar (.15 vs. .19), as indicated by another ANOVA that included study group as a factor and compared accuracy in the generate/letter height conditions against the respective nonstudied conditions. That ANOVA revealed a significant priming effect (generate/letter height vs. nonstudied), $F(1, 46) = 49.22$, $MSE = .015$, that did not interact with study group ($F < 1$). Thus, although the letter-height task did not produce quite as much priming as the read condition in this experiment (unlike Experiment 2), it did produce as much priming as the generate condition. We suspect that this pattern is a consequence of an unusually high level of priming in the read condition among subjects in the letter-height study group.

Experiment 5

We have previously shown that whereas read and generate conditions produce equivalent priming in masked word identification, the read condition produces more priming than does the generate condition in word-fragment completion (MacLeod & Masson, 1997). One reason priming induced by a generate encoding task might be larger when tested with masked word identification than with word-fragment completion pertains to the level of analysis emphasized by the test. In the word-fragment completion test, analysis may emphasize the level of individual letters because whole-word information and substantial information about interletter features is not available on this test. In the masked word identification test, whole-word information and interletter features are fully available, albeit temporarily. If we assume that the generate and letter-height encoding tasks induce subjects to process whole-word forms and interletter features, then these tasks would be particularly appropriate as a source for priming on the masked word identification task but less well suited to word-fragment completion.

In Experiment 5, we used the previously demonstrated dissociation between masked word identification and word-fragment completion tests to evaluate further our proposition that the generate and letter-height encoding tasks produce priming for the same reason: orthographic recoding. We hypothesized that if this is the case, then on a word-fragment completion test both encoding tasks should produce less priming than will a read encoding task, but, as in Experiment 4, priming in the generate and letter-height conditions should be comparable.

Method

Subjects. Forty-eight subjects from the same source as the earlier experiments were tested. Twenty-four subjects were randomly assigned to each of the two study groups.

Materials and procedure. The 90 critical targets and cues from Experiments 3 and 4 were used. For each target, a word fragment was created in which all but two or three letters were replaced by a hyphen (e.g., "*heart*"–"*h*—*r*-"). Many fragments had more than one valid completion. The study-phase procedure was the same as in Experiment 4, but in the test phase, a word-fragment completion test replaced the masked word identification test. On each test trial, subjects were shown a single word fragment and attempted to produce an oral completion for the fragment. If the subject was unable to provide a response within 15 s, the next trial began automatically. There was no titration procedure here, so the proportion of target completions in the test phase that were studied items was .67.

Results and Discussion

The mean proportions of correct study-phase responses for the generate and letter-height tasks were .93 and .95, respectively.

The mean proportions of word fragments completed with target words are shown in Figure 4. An ANOVA of these data with study group (generate, letter height) as a between-subjects factor and encoding condition (read, generate/letter height, nonstudied) as a within-subject factor found only a main effect of encoding condition, $F(2, 92) = 50.46$, $MSE = .004$. Neither the effect of study group nor the interaction approached significance ($Fs < 1$). In contrast to Experiment 4, however, pairwise comparisons collapsing across study groups and using the interaction error term from the ANOVA showed that word-fragment completion in the generate and letter-height conditions combined was lower than in the read condition ($d = .05$), $F(1, 92) = 13.69$, but that the generate and letter-height conditions combined did produce reliable priming ($d = .08$), $F(1, 92) = 38.91$.

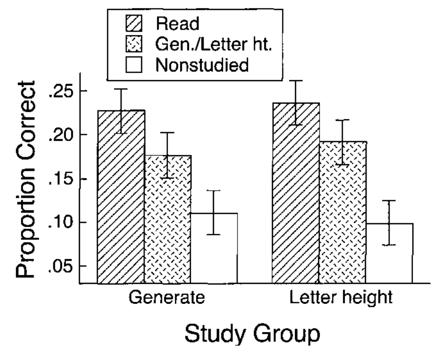

Figure 4. Experiment 5: Mean proportion correct in the word-fragment completion test as a function of study group and encoding task. Error bars show the 95% within-subject confidence intervals for each study group. Gen. = generate; Letter ht. = letter height.

This finding is consistent with results found for the generate encoding task by MacLeod and Masson (1997) and by Weldon (1991). The similarity in the pattern of priming effects for the generate and letter-height tasks relative to the read task, across both the masked word identification task of Experiment 4 and the word-fragment completion task of Experiment 5, is consistent with the hypothesis that the priming produced by these two tasks arises from a similar source, namely, covert orthographic encoding of unseen targets during the study phase.

Experiment 6

In Experiment 6, we used the letter-height task for the first time as the test to measure priming. Because of the high level of accuracy on this task, we switched to response latency as the primary dependent measure. An advantage of moving to response latency instead of accuracy is that doing so provides further generalization of priming effects produced by encoding tasks that do or do not involve direct visual encoding.

Two versions of Experiment 6 were conducted. In the first version (6A), the letter-height task was also used as one of the two encoding tasks. The other encoding task, repeat only, was the auditory task from Experiment 2 in which subjects simply repeated aloud auditorily presented targets. If the letter-height task is specifically sensitive to prior covert orthographic encoding of target words when used as an indirect test of memory, then the letter-height encoding task should produce more priming on the subsequent letter-height test task than should the repeat-only encoding task. Of course, this predicted pattern would also be expected on the basis of response priming and item-specific practice effects, given that the study and test tasks are identical in the case of the letter-height task. As in past experiments, the repeat-only condition should produce little or no priming.

The more critical version of Experiment 6 was the second one (6B). Here, a read encoding task was compared with the repeat-only encoding task. According to the orthographic-recoding hypothesis, the orthographic processing afforded by the read task should lead to more priming on the letter-height test task than should the repeat-only encoding task. This result would parallel that observed when the indirect test was masked word identification in Experiment 2. Again, repeat-only encoding was expected to yield little or no priming.

Method

Subjects. Forty-eight subjects were drawn from the same source used in the earlier experiments. Twenty-four subjects participated in each version of Experiment 6.

Materials. The critical target words were the same as those used in Experiments 3–5. An additional set of practice words was used in each version of the experiment.

Procedure. Subjects were tested using the same equipment as in the earlier experiments. In the study phase, they performed two encoding tasks: in Experiment 6A, letter height and repeat only; in Experiment 6B, read and repeat only. The encoding-task procedures were the same as in prior experiments, and four practice items were presented for each encoding task before the critical items were presented in that task. Practice trials were followed by presentation of 30 critical items in each encoding task. The two encoding tasks were presented as separate blocks of trials with task order counterbalanced across subjects. Blocked presentation of encoding tasks was used to reduce the possibility of carry-over effects across trials involving different encoding tasks that might induce subjects to engage in covert orthographic recoding of targets when they otherwise might not. Assignment of sets of 30 critical items to the two encoding tasks and to the nonstudied condition in each version of the experiment was counterbalanced across subjects so that each set of items was assigned equally often to each condition.

The test phase in each experiment began with a set of practice trials (40 in Experiment 6A and 30 in Experiment 6B) followed by 90 critical trials. Including the practice trials, the proportions of targets in the test phase that had been studied were .46 and .50 in Experiments 6A and 6B, respectively. In the letter-height test, subjects were instructed to press a button mounted on a response box as soon as they orally reported the number of ascending and descending letters in the target word. Subjects were also instructed to respond as quickly and as accurately as possible. The response box was connected to the computer keyboard, and buttonpresses were detected by the computer, permitting the measurement of response latency. Latency was measured from the offset of the auditory presentation of the target word. After each response, the experimenter pressed a key on the computer keyboard to record the subject's oral response.

Results and Discussion

Subjects responded correctly on .99 of the study trials involving the letter-height task in Experiment 6A. Outliers in the response-latency data were defined as latencies that fell outside a range that was defined so as to exclude no more than 0.5% of the trials (Ulrich & Miller, 1994). The lower bound of the cutoff for this and the remaining experiments reported here was 300 ms. The upper bound for Experiments 6A and 6B was 5,000 ms. With these cutoffs in place, the percentage of trials classified as outliers was 0.37% for both Experiments 6A and 6B. Mean response latency for trials on which a correct response was made and mean proportion error in each condition are shown in Figure 5 for each version of Experiment 6. Separate ANOVAs were computed for each experiment version, with response latency and proportion error as dependent measures and encoding condition as a repeated measures factor.

In Experiment 6A, there was a significant difference in response latency between study conditions, $F(2, 46) = 5.02$, $MSE = 7{,}423$.

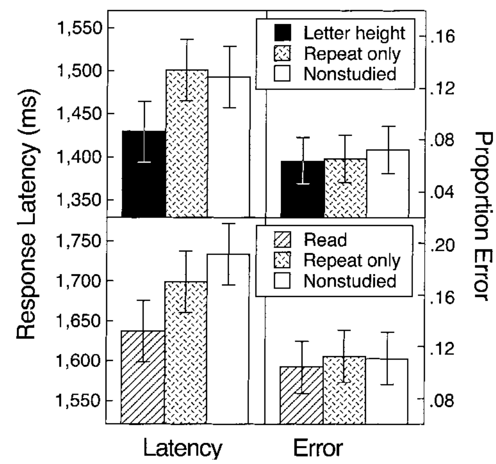

Figure 5. Experiment 6: Mean response latency and proportion error in the letter-height test in Experiments 6A (top section) and 6B (bottom section). Error bars show the 95% within-subject confidence intervals.

Planned comparisons using the error term from the ANOVA revealed that response latency in the letter-height study condition was shorter than in the repeat-only condition ($d = 72$ ms), $F(1, 46) = 8.39$, and that the repeat-only and nonstudied conditions did not reliably differ ($d = 8$ ms; $F < 1$).

In Experiment 6B, study conditions again were reliably different, $F(2, 46) = 6.58$, $MSE = 8{,}655$. Corresponding to Experiment 6A, planned comparisons indicated that the read condition led to a shorter mean latency than did the repeat-only condition ($d = 62$ ms), $F(1, 46) = 5.24$, which again did not reliably differ from the nonstudied condition ($d = 35$ ms), $F(1, 46) = 1.67$. ANOVAs were also computed in each version of the experiment using proportion error as the dependent measure. Neither of these analyses revealed a significant effect of study condition (both Fs < 1).

These results confirm each of the predictions regarding relative amounts of priming on the letter-height indirect test produced by letter-height, read, and repeat-only study tasks. As evidence that priming as measured by the letter-height indirect test is sensitive to experience with a word's orthography, the read task produced more priming than the repeat-only task. Neither Experiment 6A nor 6B produced evidence for significant priming when subjects merely heard and repeated target words during study, replicating the pattern in Experiment 2 when the indirect test was masked word identification. Combining the data from Experiments 6A and 6B also failed to produce a reliable priming effect for the repeat-only condition (1,600 vs. 1,613 ms; $F < 1$). The power of a directional version of this test to detect an effect of 50 ms (which is near the lower bound of significant priming effects seen on the letter-height test across the experiments reported in this article) was estimated to be .85. Thus, although there may be a small cross-modal priming effect on the letter-height test because of auditory study, the effect is clearly not as large as that found when subjects have either direct or internally generated experience with a word's orthography.

The pattern of priming effects produced by the letter-height and read study tasks indicates that priming on the letter-height task when used as the test is due to prior orthographic encoding experience (covert in the case of the letter-height encoding task and overt in the case of the read task) as opposed to auditory experience with the word. There may be additional sources of priming (such as response repetition) at work when the letter-height task itself is also used at encoding, but these sources were not evident in Experiment 6 because the priming effect for the read encoding task was at least as large as that for the letter-height encoding task. The weak and nonsignificant priming seen with the repeat-only study task indicates that merely hearing a target word contributes little or nothing to subsequent priming on the letter-height test.

Experiment 7

Experiment 6 provided support for the idea that priming on the letter-height test is specifically dependent on orthographic processing of targets. Therefore, priming on that test may be treated as a signature of orthographically based priming. Accordingly, in Experiment 7, priming based on read and generate study tasks was directly compared on two indirect tests: masked word identification (7A) and letter-height judgment (7B). We hypothesized that if the source of priming that arises from generation is covert orthographic processing at the time of study, then the pattern of priming produced by read and generate study tasks on the masked word identification test should be comparable with that found on the letter-height test. This outcome would support the hypothesis that priming on perceptually based indirect tests of memory that is produced by generating target words at study results from orthographic recoding that occurs in the course of generating the targets, not from the conceptual processing required by target generation.

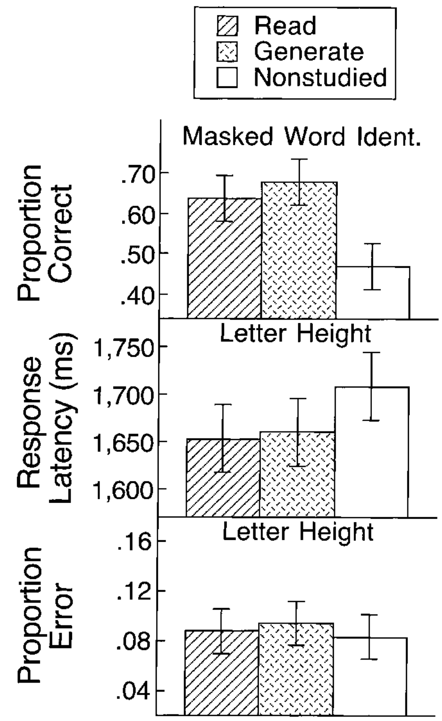

Figure 6. Experiment 7: Mean proportion correct in the masked word identification test in Experiment 7A and mean response latency and proportion error in the letter-height test in Experiment 7B. Error bars show the 95% within-subject confidence intervals. Ident. = identification task.

Method

Subjects. Forty-eight subjects from the same source as in the earlier experiments were tested, 18 in Experiment 7A (masked word identification test) and 30 in Experiment 7B (letter-height test).

Materials. The 90 critical targets and generation cues from Experiments 3–6 were used. There were also eight practice items, four for the read study task and four for the generate study task. In addition, 45 words were selected for use in a titration procedure for the masked word identification task, as in Experiments 1, 2, and 4. A set of 30 words was used as practice items at the start of the test phase for the letter-height test.

Procedure. In the study phase, the read and generate study trials were presented in separate blocks as in Experiment 6. Order of the study blocks was counterbalanced across subjects. In the test phase, subjects in Experiment 7A were given 45 titration trials, followed by 90 critical trials. Thus, including the titration trials, .44 of the trials in the test phase presented studied items. These trials were presented using the same procedure as in Experiments 1, 2, and 4. Subjects in Experiment 7B were presented 30 practice trials then 90 critical trials, following the procedure used in Experiment 6. Including the practice trials, .50 of the items presented in the test phase had been presented in the study phase.

Results and Discussion

The mean proportion of correctly generated targets in the study phase was .95 in both Experiments 7A and 7B. In the test phase, order of study block did not significantly influence performance, therefore the data were collapsed across that factor.

The mean proportions of correct masked word identification as a function of study condition in Experiment 7A are shown in the upper section of Figure 6. An ANOVA of these data, with study condition as a repeated measures factor, revealed that the study conditions produced reliably different proportions of correct responses at test, $F(2, 34) = 15.34$, $MSE = .014$. Planned comparisons using the error term from that ANOVA indicated that the read and generate conditions combined produced reliable priming relative to the nonstudied condition ($d = .19$), $F(1, 34) = 29.55$, and that the read and generate conditions did not significantly differ from each other ($d = .04$), $F(1, 34) = 1.13$.

Turning to Experiment 7B, response latencies were considered outliers and were excluded from analyses if they were less than 300 ms or longer than 4,500 ms. This constraint resulted in the exclusion of 0.48% of the observations. The mean correct response times for the letter-height test are shown in the middle section of Figure 6. An ANOVA of these data, with study condition as a repeated measures factor, indicated that the effect of study condition approached significance, $F(2, 58) = 2.81$, $MSE = 9{,}661$, $p < .07$. Planned comparisons using the error term from that analysis showed the same pattern of outcomes as on the masked word identification test. That is, the read and generate conditions combined produced a reliable priming effect ($d = 52$ ms), $F(1, 58) = 5.53$, and there was no significant difference between the read and generate conditions ($d = 7$ ms; $F < 1$). The mean proportion error on the letter-height task is shown for each study condition in the lower section of Figure 6. An ANOVA of those data indicated that the study conditions did not significantly differ with respect to proportion error ($F < 1$).

The primary outcome of Experiments 7A and 7B was a strikingly parallel effect of read and generate study conditions on two different indirect tests: masked word identification and letter height. The two study tasks—read and generate—produced equal priming on both of those tests. The results for the masked word identification test (7A) replicate the results of Experiments 1 and 4 and results reported in our earlier articles (MacLeod & Masson, 1997; Masson & MacLeod, 1992).

Experiment 8

The final experiment was designed to address two important issues. First, despite the dissociations we have established in the earlier experiments, a critic might argue that the influence of generate encoding tasks on the letter-height test could be the result of conscious recollection. Second, the evidence we have presented so far in support of the proposition that generation-based priming is a product of orthographic recoding of generated targets is somewhat indirect. In particular, we have thus far shown that generate and letter-height encoding tasks produce parallel priming effects on two different indirect tests of memory (masked word identification and word-fragment completion), and we have shown that the generate and read encoding tasks yield parallel priming effects on the letter-height test. The importance of orthographic information in both the read and the letter-height tasks suggests that the parallels seen with the generate task arise because priming caused by the generate task also depends on orthographic processing. Yet, this argument is rather indirect.

To gain further support for the proposal that orthographic recoding is the basis for priming induced by the generate task, and to obtain further evidence against the hypothesis that conscious recollection is the cause of such priming, we used a version of the generate encoding task that we expected would be unlikely to invoke orthographic recoding. Selection of this particular generate task was motivated in part by Jacoby's (1983) demonstration that generation of target words from their antonym cues produced little or no priming on the masked word identification test (see also Masson & MacLeod, 1992, Experiment 2). The lack of priming of targets generated from antonyms is an unusual outcome given the range of generate cues that successfully lead to priming on masked word identification (Masson & MacLeod, 1992). We suspect that the relatively strong association between members of antonym pairs leads to particularly fluent target generation, such that subjects may be less likely to engage in orthographic recoding as part of that generation process. Moreover, the absence of the target's first letter in the antonym-generation cues provided by Jacoby may have contributed to a generation context in which subjects were not inclined to internally construct a generated target's orthographic pattern (although Masson & MacLeod, 1992, provided the first letter of the target along with its antonym cue and still failed to obtain priming).

The version of the generate task that we adopted, then, involved presenting subjects with a target word's antonym as the cue, but without the target word's first letter. To reduce further the possibility that subjects would covertly construct orthographic codes for the target words, we presented the antonym cues auditorily. Thus, no orthographic information was available during the antonym-generation task. We predicted that under these conditions, the antonym-generation task would yield little or no priming on the letter-height test, which we have shown to be specifically sensitive to prior orthographic experience (Experiment 5). This result would replicate the findings obtained by Jacoby (1983) and Masson and MacLeod (1992, Experiment 2) for the masked word identification test.

In contrast to the antonym-generate encoding task, we expected that a read encoding task would once again lead to priming on the letter-height test. In addition, to establish a clear double dissociation between the letter-height test and a direct test of memory, a second version of Experiment 8 was conducted in which a recognition test was administered rather than the letter-height test. In this case, we expected that the generate task would lead to much better memory than the read task, thereby establishing a double dissociation between the letter-height and recognition tests. That dissociation would support the claim that priming on the letter-height test has little or no grounding in conscious recollection.

Method

Subjects. Thirty subjects from the same source as the previous experiments took part in Experiment 8; 24 in Experiment 8A (letter-height test) and 6 in Experiment 8B (recognition).

Materials. A set of 81 pairs of antonyms was selected as critical items. The target member of each pair was arbitrarily selected, except that it was required to be at least four letters in length. The normative frequency of occurrence of the targets ranged from 0 to 895 per million, with a median of 88 (Kučera & Francis, 1967). Six additional pairs were selected for use as practice items in the study phase. and four other words were chosen for use as practice items in the letter-height test of Experiment 8A. The 81 critical pairs were arranged as three lists of 27 items each, and assignment of these lists to study conditions (read, generate, and nonstudied) was counterbalanced across subjects so that each list was used equally often in each condition.

Procedure. The read and generate tasks were presented in blocked fashion, as in Experiment 7. Again, the expectation was that with blocked presentation of tasks there would be less likelihood of carryover of orthographic processing across trials (in particular, from read to generate trials). The procedure for the study phase was similar to that of Experiment 7, except that the antonym cues were recorded in a male voice and presented auditorily without the first letter of the target word. The test phase for Experiment 8A (letter-height test) was similar to that of Experiment 7, except that there were only four practice trials, so that the overall proportion of test items that had been presented in the study phase was .64. The test phase for Experiment 8B (recognition test) was similar to that for Experiment 3.

Results and Discussion

The mean proportion of generate targets correctly produced was .89 in both Experiments 8A and 8B. Response latencies in the letter-height test of Experiment 8A that were less than 300 ms or longer than 5,000 ms were excluded as outliers, thereby removing 0.46% of the observations. The means for correct response latency and for error proportion in Experiment 8A are shown in Figure 7. An ANOVA with study condition as a factor was computed for the latency data. The effect of study condition approached significance, $F(2, 46) = 2.65$, $MSE = 21{,}697$, $p < .09$. Planned comparisons using the error term from that ANOVA showed that latencies in the read condition were reliably shorter than in the generate and nonstudied conditions combined ($d = 85$ ms), $F(1, 46) = 5.29$, and that the generate and nonstudied conditions did not significantly differ ($d = 4$ ms; $F < 1$). Thus, although the read condition led to priming on the letter-height test,

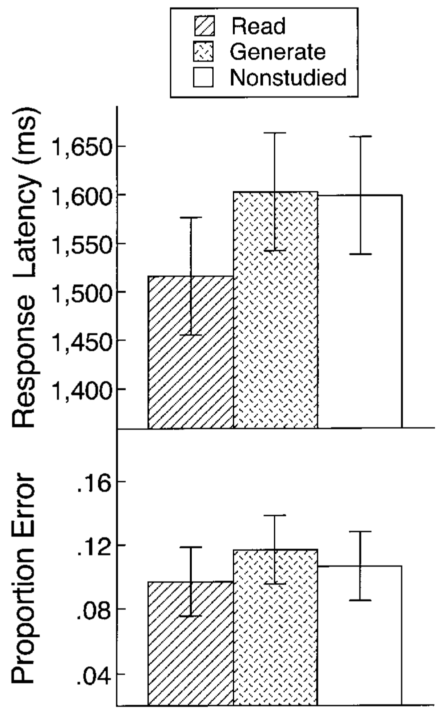

Figure 7. Experiment 8: Mean proportion correct in the masked word identification test. Error bars show the 95% within-subject confidence intervals for each study group.

the antonym-generation task did not. The mean error proportion did not differ significantly across study conditions ($F < 1$).

In Experiment 8B, the mean proportion of correctly recognized items was .41 ($SD = .20$) for the read condition and .80 ($SD = .09$) for the generate condition. The advantage for the generate condition was significant, $F(1, 5) = 20.85$, $MSE = .022$. When recognition in the generate condition was conditionalized on successful study-phase generation, the mean proportion of hits increased to .88 ($SD = .07$), making this advantage even larger. The proportion of false alarms was .08 ($SD = .04$).

These results establish the predicted double dissociation between the letter-height and recognition tests. One implication of this dissociation is that it is unlikely that conscious recollection was the basis for priming seen on the letter-height test. Awareness of prior presentation was clearly greater in the generate condition, yet this condition yielded no priming whatsoever on the letter-height test. Moreover, the proportion of test items that had been presented in the study phase was at least as high for the indirect test of memory in this experiment as in any of the other experiments reported here, yet priming was restricted to the read study condition. These facts support the conclusion that priming on the letter-height test is influenced very little, if at all, by conscious recollection.

The other important aspect of these results is that by reducing the likelihood that subjects would engage in covert orthographic recoding of targets during generation (by using auditory cues and not presenting the first letter of the target), we eliminated priming on the letter-height test for items that had been generated at study. This result suggests that the extent to which a generate encoding task invites orthographic recoding of generated targets powerfully modulates priming on perceptually based indirect tests of memory.

General Discussion

The primary motivation for the experiments reported here was to examine the basis for repetition priming on what are assumed to be perceptually based indirect tests of memory. In particular, our interest was in priming produced by a study task—generation from a semantic cue—usually regarded as nonperceptual. In our earlier work, we have consistently found that generation from a semantic cue can yield as much priming as does reading a target word during study (Bodner et al., 2000; MacLeod & Masson, 1997, 2000; Masson & MacLeod, 1992, 1997). Because such robust priming from a generation task would be unexpected in the context of the transfer-appropriate processing framework (e.g., Jacoby, 1983; Morris et al., 1977; Roediger et al., 1989), we have suggested in a series of studies over the past decade that indirect tests such as masked word identification may have a conceptual component that benefits from prior generation experience.

The alternative explanation pursued here is that priming on such indirect tests is driven by prior orthographic processing, whether directly experienced, as in the read task, or indirectly experienced, as through visualization of a word's orthography. We have argued that such orthographic recoding is a normal consequence of generation at study, though not an inevitable one (see Experiment 8; Jacoby, 1983; Masson & MacLeod, 1992, Experiment 2). Recoding likely occurs as a means of checking that the correct target word has been generated. As Farah (1988; Farah et al., 1988) and Kosslyn et al. (1993) have argued, imagining a visual stimulus can closely approximate actual visual experience of that stimulus.

In Experiment 1, we showed that the modality in which generation cues are presented (visual vs. auditory) does not influence the amount of priming obtained in a subsequent masked word identification test. This outcome helped to set the stage for a comparison between the generate task (in which generation cues typically are presented visually) and the letter-height task that we introduced in this article, in which targets are necessarily presented auditorily. Experiment 2 provided evidence that the letter-height task produced priming on the masked word identification test comparable with that found when the encoding task involved reading, whereas merely hearing and repeating target words failed to lead to priming. This result supported the hypothesis that the letter-height task induced a form of orthographic recoding that could support priming on a perceptually based indirect test of memory.

Experiment 3 simultaneously addressed the processing carried out in the generation and letter-height tasks and the possibility of intrusion from conscious recollection. Here, we used a direct test of memory and showed that recognition of words following generation at study was far superior to recognition of words following letter-height judgment at study. This confirms the strong conceptual processing engaged by the generate task and suggests that the letter-height task emphasizes the physical (i.e., nonconceptual) structure of the target words. Were conscious intrusion operating, then, the generate task would be expected to lead to a considerably better performance than would the letter-height task. Once again, our results are inconsistent with such a pattern and hence with the conscious-intrusion explanation.

In Experiments 4 and 5, the generate task and the letter-height task produced, to a large extent, parallel patterns of priming on two different indirect tests, masked word identification and word-fragment completion. On the masked word identification test, both encoding tasks produced similar amounts of priming that were, in turn, comparable with priming found with the read encoding task (although there was a trend toward less priming in the letter-height condition than in the read condition). On the word-fragment completion test, in contrast, there was significantly less priming for both the generate and the letter-height tasks than for the read task, although equal priming was again found for the generate and letter-height tasks. This pattern across these two tests is entirely consistent with our previous work comparing these two forms of indirect test (MacLeod & Masson, 1997).

In Experiment 6, we established that the letter-height task could be used as an indirect test of memory that is specifically sensitive to prior orthographic experience with target words. Both read and letter-height encoding tasks produced reliable priming on this test, whereas merely hearing and repeating targets led to little or no priming. Then, in Experiment 7, we demonstrated that read and generate encoding tasks led to comparable amounts of priming on both the masked word identification test and the letter-height test, suggesting that both encoding tasks produced priming because of the orthographic experience that they afford—directly in one case and covertly in the other.

Finally, in Experiment 8, we took advantage of the previously reported finding (Jacoby, 1983; Masson & MacLeod, 1992) that generating from an antonym does not produce reliable priming on a masked word identification test. This may be because the task of generating antonyms is sufficiently well defined and easy, so that checking the generated target, and hence orthographically recoding the target, is not necessary and therefore is not carried out. We therefore predicted that there would be no reliable priming on the letter-height test following antonym generation, and that is what we found, confirming again the perceptual nature of the letter-height task whether at study or at test. We also showed that the antonym-generation task led to much higher recognition memory than did the read task, replicating Jacoby's (1983) and Masson and MacLeod's (1992) studies and ruling out conscious recollection as a possible source of priming on the letter-height test.

Our earlier work has shown that generation of targets from semantically related phrases can lead to priming in masked word identification, word-fragment completion, word-stem completion, and speeded word reading (Bodner et al., 2000; MacLeod & Daniels, 2000; MacLeod & Masson, 1997, 2000; Masson & MacLeod, 1992). We have previously advocated a conceptual basis for generation-based priming in these tests because generation certainly requires conceptual processing and certainly does not provide direct perceptual experience with the target (e.g., Masson & MacLeod, 1997). Yet these indirect tests have usually been characterized as data-driven, so why should they reveal a benefit from conceptual encoding?

The results reported here establish that a viable alternative explanation of those priming effects in fact does have a data-driven rather than conceptual basis. This alternative explanation fits with the transfer-appropriate processing account of priming effects (Jacoby, 1983; Roediger et al., 1989). In that account, data-driven processes engaged during study are likely to be recapitulated if the test task requires data-driven processing of a previously studied target, leading to a priming effect. These results are also compatible with Bowers's (2000) proposal that repetition priming in word identification tasks is based on improved orthographic encoding of targets. It is intriguing to note that the present study suggests strongly that data-driven encoding does not require the physical stimulus but instead functions almost equally effectively with a visualization of that stimulus.

The orthographic-recoding hypothesis we have examined fits with a number of other results regarding priming on indirect data-driven tests. First, auditory encoding tasks produce little or no priming on masked word identification, as shown here (Experiments 2 and 6) and elsewhere (e.g., Jacoby & Dallas, 1981; Kirsner, Milech, & Standen, 1983; Weldon, 1991). Auditory encoding tasks would not ordinarily be expected to induce covert orthographic processing of targets. Evidence supporting the occurrence of orthographic recoding of an auditorily presented word is based on encoding tasks that invite subjects to analyze a word's phonemic components (Neaderhiser & Church, 2000; Seidenberg & Tanenhaus, 1979), an analysis that is unlikely to occur if subjects are merely required to repeat an auditorily presented word. Demonstrations of small cross-modal priming effects from auditory study to a visual indirect test of memory such as word-fragment completion (e.g., Roediger & Blaxton, 1987; Weldon, 1991) could be the result of occasional occurrences of orthographic recoding.

Second, it is curious that the generation and letter-height encoding tasks produced about as much priming as the read encoding task on some indirect tests (masked word identification and letter-height tests) but not on others (word-fragment completion). We suspect that an important distinction between these two situations hinges on whether the test requires an internal regeneration of a target word's orthography, as in the masked word identification and letter-height tests. In those cases, the subject either is presented with only a brief glimpse of the target or only hears the target and must generate the target in a designated format from that event. This regeneration process may be a particularly strong fit for the prior encoding task of either generating the target from a semantic cue or generating the target's orthography from an auditory presentation. When tested with a word fragment, the available letters are constantly in view, and the subject must regenerate the remaining letters. Therefore the word-fragment completion test may not require the same kind of regeneration as that demanded by masked or auditory presentations. Thus, direct perceptual experience may have an advantage over comparable imagined experience (e.g., McDermott & Roediger, 1994; Pilotti, Gallo, & Roediger, 2000), but this advantage may be overcome when the test demands fit particularly well with a generation process invoked by the encoding task.

An exception to this proposal about target regeneration is our earlier finding (MacLeod & Masson, 2000) that generate encoding tasks led to about as much priming as the read encoding task when the subsequent indirect test involved speeded reading of word targets. In that test, the targets were clearly visible and would not have required the kind of regeneration presumably demanded by masked word identification and letter-height tests. We note, however, that there was a consistent tendency in that study for the generate task to produce somewhat less, though not significantly less, priming than the read task. It is quite possible that a more powerful test would have revealed reliably more priming from the read task than from the generate task, as in fact was reported with a considerably larger sample by MacLeod and Daniels (2000).

Although our results do not rule out the possibility that conceptual encoding processes can induce priming on perceptually based indirect tests of memory such as masked word identification, they suggest that priming on such tests may be affected very little or not at all by conceptual aspects of encoding episodes. There are conditions, however, under which conceptual encoding operations can influence later word identification. Toth and Hunt (1990) found an advantage for generate over read items on a masked word identification test if targets were both studied and tested in the presence of an associatively related cue word. This effect was not likely due to conscious recollection strategies because it was replicated when subjects were tested under pharmacologically induced amnesia (Hirshman et al., 2001).

We suggest that by including an associative cue at test, Toth and Hunt (1990) transformed the word identification task into a conceptually oriented one that arguably is rather different from the version used here and in most previous studies. As evidence of this difference, we note that when Toth and Hunt used test-phase catch trials on which a previously studied cue was followed by a related but nonstudied target word that had to be identified, subjects responded (erroneously) on over one third of these trials with the item that had been studied with the cue. Thus, when faced with impoverished perceptual evidence, subjects often resorted to a conceptually based response. It is also interesting to note that these intrusions were more likely to occur in response to cues that had appeared at study with generated targets rather than with read targets. Moreover, the size of this difference in intrusion rates was at least as large as the advantage of generate over read items in correct identification. Thus, the generation effect reported by Toth and Hunt appears to be driven by conceptual processes enabled by the presentation of associative cues at the time of test. Our evidence suggests that in the absence of these cues, tests such as masked word identification leave little room for conceptually based influences of prior study episodes. One might reasonably suppose that in natural reading situations, contextual cues in the form of neighboring words typically contribute to word identification. These circumstances, however, appear to involve a set of constraints that are not operative when identifying words presented in isolation (e.g., Levy & Kirsner, 1989; MacLeod, 1989; Masson & MacLeod, 2000).

The findings of the present experiments provide significant evidence in support of an orthographic source, rather than a conceptual source, for generation-based priming. In the past (e.g., Masson & MacLeod, 1992), we had argued that because generation produces priming that could not be due to actual perceptual experience, then such priming is evidence of a conceptual contribution to priming. Our position has changed. We now consider the orthographic-recoding account to be a viable explanation of generation-based priming on indirect tests of memory that primarily depend on data-driven processes. This explanation is reminiscent of the idea that imaging uses many of the same processes as does actually perceiving. Thus, an imagined orthography can constitute the data for a data-driven memory test. In consequence, we interpret our results as being entirely consistent with the transfer-appropriate processing framework.

References

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 657–668.

Bodner, G. E., Masson, M. E. J., & Caldwell, J. I. (2000). Evidence for a generate–recognize model of episodic influences on word-stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 267–293.

Bowers, J. S. (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 7,* 83–99.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307–317.

Farah, M. J., Peronnet, F., Gonon, M. A., & Girard, M. H. (1988). Electrophysiological evidence for a shared representational medium for visual images and visual percepts. *Journal of Experimental Psychology: General, 117,* 248–257.

Hirshman, E., Passannante, A., & Arndt, J. (2001). Midazolam amnesia and conceptual processing in implicit memory. *Journal of Experimental Psychology: General, 130,* 453–465.

Horton, K. D., Wilson, D. E., & Evans, M. (2001). Measuring automatic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 958–966.

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., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory. *Journal of Experimental Psychology: General, 122,* 139–154.

Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36,* 300–324.

Kirsner, K., Milech, D., & Standen, P. (1983). Common and modality-specific processes in the mental lexicon. *Memory & Cognition, 11,* 621–630.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., et al. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience, 5,* 263–287.

Kučera, J., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Levy, B. A., & Kirsner, K. (1989). Reprocessing text: Indirect measures of word and message level processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 407–417.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

MacLeod, C. M. (1989). Word context during initial exposure influences degree of priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 398–406.

MacLeod, C. M., & Daniels, K. A. (2000). Direct versus indirect tests of memory: Directed forgetting meets the generation effect. *Psychonomic Bulletin & Review, 7,* 354–359.

MacLeod, C. M., & Masson, M. E. J. (1997). Priming patterns are different in masked word identification and word fragment completion. *Journal of Memory and Language, 36,* 461–483.

MacLeod, C. M., & Masson, M. E. J. (2000). Repetition priming in speeded word reading: Contributions of perceptual and conceptual processing episodes. *Journal of Memory and Language, 42,* 208–228.

Masson, M. E. J. (in press). Bias effects in repetition priming of masked word identification: Differential influences of modality and generation. *Psychonomic Bulletin & Review.*

Masson, M. E. J., & MacLeod, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General, 121,* 145–176.
Masson, M. E. J., & MacLeod, C. M. (1997). Episodic enhancement of processing fluency. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 37, pp. 155–210). San Diego: Academic Press.
Masson, M. E. J., & MacLeod, C. M. (2000). Taking the "text" out of context effects in repetition priming of word identification. *Memory & Cognition, 28,* 1090–1097.
McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review, 4,* 582–586.
McDermott, K. B., & Roediger, H. L., III. (1994). Effects of imagery on perceptual implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1379–1390.
McKenna, P., & Warrington, E. K. (1993). The neuropsychology of semantic memory. In I. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 193–213). Amsterdam: Elsevier.
McKone, E., & Murphy, B. (2000). Implicit false memory: Effects of modality and multiple study presentations on long-lived semantic priming. *Journal of Memory and Language, 43,* 89–109.
Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 15,* 519–533.
Neaderhiser, B. J., & Church, B. A. (2000, November). *Priming perceptually driven tasks without a percept*. Paper presented at the 41st Annual Meeting of the Psychonomic Society, New Orleans, LA.
Pentland, A. (1980). Maximum likelihood estimation: The best PEST. *Perception & Psychophysics, 28,* 377–379.
Pilotti, M., Gallo, D. A., & Roediger, H. L., III (2000). Effects of hearing words, imagining hearing words, and reading on auditory implicit and explicit memory tests. *Memory & Cognition, 28,* 1406–1418.
Rajaram, S., & Roediger, H. L., III. (1993). Direct comparison of four implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19,* 765–776.
Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review, 104,* 319–343.
Roediger, H. L., III. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45,* 1043–1056.
Roediger, H. L., III, & Blaxton, T. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & Cognition, 15,* 379–388.
Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.
Roediger, H. L., III, Weldon, M. S., Stadler, M. L., & Riegler, G. L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1251–1269.
Schacter, D. L., & Graf, P. (1989). Modality specificity of implicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 3–12.
Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 546–554.
Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4,* 592–604.
Toth, J. P., & Hunt, R. R. (1990). Effect of generation on a word-identification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 993–1003.
Toth, J. P., Reingold, E. M., & Jacoby, L. L. (1994). Toward a redefinition of implicit memory: Process dissociation following elaborative processing and self-generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 290–303.
Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123,* 34–80.
Weldon, M. S. (1991). Mechanisms underlying priming on perceptual tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 526–541.

Received July 20, 2001
Revision received March 1, 2002
Accepted March 20, 2002 ■