

## Modality Effects in False Recall and False Recognition

David A. Gallo and Kathleen B. McDermott  
Washington University in St. Louis

Jenny M. Percer  
Stanford University

Henry L. Roediger III  
Washington University in St. Louis

R. E. Smith and R. R. Hunt (1998) reported a dramatic reduction in false remembering in a list-learning paradigm by switching from auditory to visual presentation at study. The current authors replicated these modality effects using written recall and visual recognition tests but obtained smaller effects than those in R. E. Smith and R. R. Hunt's study. In contrast, no modality effect occurred on auditory recognition tests. Manipulating study and test modality within-subjects (Experiment 2) and between-subjects (Experiment 3) yielded similar results. It was also found that subjects frequently judged critical nonstudied words as having been presented in the modality of their corresponding study lists. The authors concluded that subjects could retrieve distinctive information about a study list's presentation modality to reduce false remembering but only did so under certain conditions. The modality effect on false remembering is a function of both encoding and retrieval factors.

The Deese–Roediger–McDermott (DRM) list-learning paradigm provides a tractable means by which false memories can be created and studied in the laboratory (Deese, 1959; Roediger & McDermott, 1995). Roediger and McDermott (1995) presented subjects with lists of words for immediate free recall. Unbeknownst to the subjects, each list consisted of 15 items (e.g., *bed, rest, awake*) that were the most common responses to a critical item (e.g., *sleep*) in free association norms. Even though this critical item was not included in the study list, subjects erroneously recalled it with approximately the same probability as they recalled studied items from the middle of the list. On a final recognition test, administered after multiple lists had been studied and recalled, the false-alarm rate to critical items was as high as the hit rate to studied items. In addition, false recognition was accompanied by high confidence ratings and, in another experiment, subjects claimed to “remember” the occurrence of the critical item in the study list as often as they “remembered” list items (using Tulving's, 1985, “remember”/“know” procedure). These results have been replicated and extended by many subsequent researchers (see Roediger, McDermott, & Robinson, 1998, for a partial review). One reason for the popularity of this paradigm is that, much like a perceptual illusion, this memory illusion is compelling and diffi-

cult to avoid. Even subjects who are explicitly warned against making such memory errors cannot eliminate false remembering (Gallo, Roberts, & Seamon, 1997; McDermott & Roediger, 1998).

Exactly how these false memories are created and their implication for theoretical accounts of memory are topics of considerable debate. One theory posits that the critical item becomes activated as an implicit associate response to its associates in the study list (Underwood, 1965). This activation could result from automatic spreading activation within an associative network (e.g., Roediger, Balota, & Watson, 2001; Seamon, Luo, & Gallo, 1998), or from conscious thought of the critical item at study (e.g., McDermott, 1997), or both. Once the critical item has been activated, subjects later falsely recall or recognize this item because of a failure to monitor correctly the source of this item's activation (i.e., a failure of reality monitoring; Johnson & Raye, 1981). Others have argued that these false memories are caused by the encoding of a general sense of the thematic content of the list rather than the critical item itself. For instance, fuzzy-trace theory stipulates that subjects encode both specific (verbatim) and general semantic (gist) information about events (Reyna & Brainerd, 1995). False recall and recognition could be caused by the match between the critical item and the gist representation of the list, which would suggest to the subject that this item had been studied (e.g., Payne, Elie, Blackwell, & Neuschatz, 1996). If the gist representation is sufficiently strong, then the critical item may elicit the illusory subjective experience of its actual presentation (i.e. “phantom recollection,” see Brainerd, Wright, Reyna, & Mo-jardin, 2001).

One theoretically important issue is that certain encoding manipulations may lead to reductions in false memories through metamemorial processes occurring at retrieval. For instance, several researchers have demonstrated that slowing presentation rate decreases the probability of false remembering (e.g., Gallo and Roediger, 2000; McDermott & Watson, in press; Togliola & Neu-

---

David A. Gallo, Kathleen B. McDermott, and Henry L. Roediger III, Department of Psychology, Washington University in St. Louis; Jenny M. Percer, Department of Psychology, Stanford University.

We thank Elizabeth Marsh for thoughtful discussions and comments, and also Charles Brainerd, Reed Hunt, and Stephen Lindsay for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to David A. Gallo, Department of Psychology, Washington University in St. Louis, Campus Box 1125, One Brookings Drive, St. Louis, Missouri 63130. Electronic mail may be sent to dgallo@artsci.wustl.edu.

schatz, 1996; but see Arndt & Hirshman, 1998). From a purely encoding viewpoint these results are paradoxical. If anything, slower presentation rates should result in either more activation of the critical item or a stronger gist representation, thereby resulting in more false remembering. Further, providing subjects with more time to encode list items should offer more opportunity for "deep" or meaningful processing of the list items, which has been found to enhance the false memory effect (e.g., Rhodes & Anastasi, 2000; Thapar & McDermott, in press; Toggia, Neuschatz, & Goodwin, 1999). The fact that false remembering is reduced at slower presentation rates may occur because at slower rates, each list item could be processed more extensively, thereby resulting in more item-specific information. At retrieval, this additional information would enhance the discriminability between those items that were presented and those that were not, resulting in more accurate reality monitoring (or more reliance on verbatim information) and less false remembering.<sup>1</sup>

In a similar vein, Israel and Schacter (1997) argued that subjects could reduce false remembering at retrieval by using the distinctiveness of the format in which studied items were presented. They presented subjects with lists comprising the top 12 semantic associates to critical nonpresented items, with the provision that all items could be pictorially represented. For one group of subjects, study words were simultaneously presented auditorily and visually (the word-only condition). For another group, study words were simultaneously presented auditorily and as black and white line drawings (the picture + word condition). False recognition of the critical items, tested after all the lists had been presented, was considerably lower for subjects in the picture + word condition than in the word-only condition.

Israel and Schacter (1997) argued that the study of pictures resulted in more perceptually distinct memories than did the study of words. As a result, subjects in the picture + word condition may have used a distinctiveness heuristic at retrieval to reduce false recognition. As defined by Israel and Schacter, the *distinctiveness heuristic* is a metamemorial process that subjects may use at the time of retrieval to help decide whether a test item had been studied (see Strack & Bless, 1994, for a discussion of such processes). Under conditions in which a subject believes that the retrieval of a studied item should be accompanied by distinct recollections of its original occurrence, subjects may require such distinct recollections before claiming to remember an item. If such distinct information were absent, then it would suggest to the subject that the test item had not been studied. Israel and Schacter argued that subjects who had studied pictures with the words relied on the heuristic that if a test item did not bring to mind a distinct pictorial recollection, it probably had not been presented at study. These subjects could reduce false alarms to critical items because these items would be unlikely to bring to mind pictorial representations of their original occurrence. Subjects who had studied only words, however, would not demand such distinct recollections because words alone are not as perceptually distinct as pictures.

Schacter, Israel, and Racine (1999) distinguished between reductions in false remembering that result from a distinctiveness heuristic and those that result from accessing list-specific information (pp. 18–19). Whereas the former process is based on global information about the distinctiveness of all studied material, the latter is based on the relative distinctiveness between memories for specific items from a particular list and false memories for the

nonpresented critical item. Access to local (list-specific) information could reduce false remembering to the extent that memories for list items are more distinct, and hence more discriminable, than memories for critical items (as may be the case when study presentation rate is slowed). In contrast to accessing list-specific information, a distinctiveness heuristic implies a global change in response criterion based on the expected memorability of all presented items. One testable difference between these two processes is that reductions based on list-specific information should be insensitive to between-subject versus within-subject manipulations of study list distinctiveness, as long as subjects can access list-specific information in each design. In contrast, a general distinctiveness heuristic is rendered ineffective when distinctiveness is manipulated within-subjects, because under these conditions subjects cannot be sure that all studied material should bring to mind distinct recollections. We return to this distinction later in Experiment 3.

Whereas Israel and Schacter (1997) used pictorial materials to reduce false remembering in the DRM paradigm, Smith and Hunt (1998) argued that simply presenting words visually during study is sufficient to reduce false remembering, relative to the standard condition in which study words are presented auditorily (e.g., Roediger & McDermott, 1995). Smith and Hunt found dramatic reductions in false recall and false recognition, on the order of 50%, by merely switching presentation modality from auditory to visual (manipulated as a between-subjects variable). For instance, in their second experiment, the probability of false recall was .42 following auditory presentation and .22 following visual presentation, and false recognition rates on a visual test dropped from .82 to .45, respectively. They proposed that visual presentation led to more distinctive item-specific processing than auditory presentation. Thus, at retrieval, the discriminability between visual events and internally generated events would be greater than the discriminability between auditory events and internally generated events. As a result, subjects who had studied lists visually could more readily discriminate between items actually studied and those that were only thought of during study (i.e., the critical items). An alternative account of these results, as discussed by Schacter et al. (1999, p. 19), is that visual presentation reduced false remembering due to a global distinctiveness heuristic, as opposed to access to list or item-specific information. Because Smith and Hunt (1998) manipulated study modality between-subjects, both accounts are theoretically possible.

Although Smith and Hunt's (1998) study was the first one published to directly compare auditory and visual presentation in the DRM paradigm, other studies have used visual presentation alone. Robinson and Roediger (1997; see also McDermott, 1997) reported lower levels of false recall following visual presentation than that found in the comparable auditory conditions of Roediger and McDermott (1995). As Smith and Hunt pointed out, this outcome suggests that false recall is sensitive to study modality.

<sup>1</sup> We assume that these retrieval processes predominantly occur at test, when the subject would be in a retrieval mode or mental set, although it is also possible that they could occur as the list is being presented during the study phase. For instance, the subject may think of the critical item at study, but during rehearsal of the list he or she may decide that it had not been presented (before the recognition test is actually administered).

Further, several unpublished studies have evaluated the effects of study modality on false recall, and these too have demonstrated that auditory presentation leads to greater false recall than does visual presentation (Kellogg, in press; Shobe & Kihlstrom, 1999; Smith, 2000). Considered as a whole, these studies indicate that the modality effect on false recall is quite reliable.

There is less evidence in the literature suggesting that there is a dramatic reduction in false recognition following visual presentation. Indeed, several studies have found robust levels of false recognition, approximating the hit rate, following visual presentation (e.g., Arndt & Hirshman, 1998; Seamon et al., 1998). Miller and Gazzaniga (1998) reported no differences in false recognition between a condition in which lists were studied auditorily and tested visually and one in which lists were studied visually and tested auditorily, although this comparison was not the main focus of that article (see their Footnote 1). Similarly, Israel and Schacter (1997) found that false recognition did not differ between the word-only condition (items were presented both auditorily and visually at study) and a control condition (items were simply presented auditorily at study).

In a direct test of modality effects on false recognition, Maylor and Mo (1999) manipulated both study and test modality within-subjects, with an immediate 12-item recognition test after each DRM list. Maylor and Mo also found dramatic reductions in false recognition, on the order of 50%, but their effects were in the opposite direction as those found by Smith and Hunt (1998). On both auditory and visual recognition tests, Maylor and Mo found that visual study led to greater false recognition than auditory study. They argued that these results were consistent with the notion that auditory presentation led to more perceptually detailed and temporally distinct memories than visual presentation. As a result, subjects were better able to use reality monitoring to reduce false recognition following auditory study than following visual study.

The findings of Smith and Hunt (1998) and Maylor and Mo (1999) are clearly at odds. Whereas the former found that false recognition was lower following visual study, the latter found that false recognition was lower following auditory study, and both argued that the modality that caused the least false remembering was more distinct. Although it is unclear whether the auditory or visual modality leads to more distinct memories in some absolute sense, Smith and Hunt's interpretation does converge with theoretical mechanisms that may be driving false recognition. If it is assumed that at least part of the false recognition effect is due to thoughts of the critical item at study, then it is not too unreasonable to think that visual presentation (which might activate both orthographic and phonological codes) would be more discriminable from such thoughts than auditory presentation (which might activate only phonological codes). This account is also parsimonious, in that it meshes well with the finding that visual presentation leads to less false remembering than auditory presentation on free recall tests.

It is unclear why Maylor and Mo (1999) found an opposite effect of modality, but two pieces of evidence suggest that it was because the particular conditions of their experiment (ones not present in other research) led to more distinctive processing during auditory than visual study. First, recognition of list items was considerably lower following visual study ( $M = .51$ , averaging across test modality) than following auditory study (.70). Second,

subjects could see the experimenter's face as the list was read to them in the auditory condition (E. A. Maylor, personal communication, January 17, 2000). Seeing the experimenter articulate study words would have provided more perceptual information during the encoding of auditory lists, and this additional information may have made this mode of presentation more distinct than the other (in which the items were printed on paper and individually revealed by the experimenter). Of course, these accounts are speculative, and further work is needed to test the generalizability of these findings.

The modality effects reported by Smith and Hunt (1998) and Maylor and Mo (1999) deserve more careful scrutiny. The DRM paradigm is becoming widely used for studying false memories. If simply changing the modality causes large reductions in false remembering, then this finding would be methodologically and theoretically important. In our experiments we set out to further evaluate the role of presentation modality (auditory vs. visual) in the DRM paradigm. In Experiment 1, conducted shortly after Roediger and McDermott's (1995) report and well before the recent flurry of work on modality, we manipulated study modality within-subjects and tested memory using free recall and recognition tests. All other aspects of the procedure were modeled after those of Roediger and McDermott (Experiment 2). To anticipate, we found that visual presentation reduced false recall relative to auditory presentation, and there was a small effect in the same direction in false recognition. We also administered a modality-judgment test and found that subjects erroneously claimed to remember the modality in which a critical item had been presented during study. In Experiments 2 and 3 we further explored the possibility that presentation modality could affect false recognition by manipulating both study and test modality and by using either a within-subjects (Experiment 2) or a between-subjects (Experiment 3) design. By comparing these two designs, we hoped to gain leverage on whether modality effects on false recognition are due to a global distinctiveness heuristic, access to list-specific information, or both.

## Experiment 1

### Method

**Subjects.** Fifty-two Rice University undergraduates participated in fulfillment of a course requirement.

**Materials.** We used the 24 lists from the appendix of Roediger and McDermott's (1995) article with the exception of the *soft*, *girl*, and *music* lists, which were replaced with the *lion*, *trash*, and *smoke* lists reported in Stadler, Roediger, and McDermott (1999). This step was taken because the *soft* list may have been confused with the *rough* list that was used, and because the *girl* and *music* lists had been found to elicit low levels of false recall relative to other lists. Like the Roediger and McDermott lists, these new lists were created from Russell and Jenkins's (1954) word association norms. Each list consisted of the 15 most common associates to a critical item, and none of the critical items occurred in any of the lists. Words within a list were arranged in descending order of associative strength to the critical item. Study items were presented visually (in lower case) in the center of the computer screen or auditorily (in a female voice) over headphones.

**Design.** Study modality and task were manipulated within-subjects, yielding a 2 (study modality: auditory or visual)  $\times$  2 (task: recall or math after list presentation)  $\times$  2 (item type: list item or critical nonpresented item) design. The 24 study lists were presented to each subject in the same

random order, with 12 presented visually and 12 presented auditorily. Half of the lists presented in each modality were followed by an immediate free recall test, and half were followed by math problems. Thus, there was a set of six lists in each of the four modality-task conditions (auditory-recall, auditory-math, visual-recall, and visual-math). Four counterbalancing groups were created so that each list set appeared once in each modality-task condition. Presentation modality and task were mixed so subjects could not anticipate the modality of an upcoming list or whether they would recall that list or solve math problems.

*Procedure.* Subjects were tested individually by computer and wore headphones throughout the first part of the session. They were told that the experiment was designed to test their memory and math abilities and that they would be presented with several 15-word lists to remember. All study-words were presented at the rate of one word per second. Subjects were told that they would see some lists and hear others, and after each list they would either recall the list or do math problems when prompted by the words *recall* or *math*, which were spoken through the headphones. Because they could not anticipate which lists were to be recalled, it was stressed that they should always try to remember the words in each list. They were given two response booklets, one with several sheets of math problems, and one with lined sheets for recall.

Following the recall prompt, subjects wrote down, in any order, as many words as they could remember from the list that had just been presented. They were told to avoid guessing and to write down only those words that they were confident had appeared in the list. Following the math prompt, they solved as many math problems as they could within the allotted time. Subjects were given 2 min for both recall and math problems. When the 2 min had elapsed, the subjects heard "Get ready for the next list," at which time they turned to a new sheet in their response booklet and looked to the monitor (so as not to miss any of the visually presented lists).

After the final recall or math period, subjects removed their headphones and were given the "old"/"new" recognition test. The test sheets consisted of 144 randomly arranged items, half of which had been presented and half of which had not.<sup>2</sup> The studied items were sampled from Serial Positions 3, 8, and 10 of each list. The nonstudied items were the 24 critical items in addition to 48 distractors that were unrelated to the study lists. Next to each word were the words "old" and "new," as well as the letters "R" and "K" (for the "remember"/"know" judgment). Subjects were told that the test consisted of some words they had studied (old) and some words they had not (new). They were instructed to circle "old" if they were confident that they had seen or heard the word at study. If they did not think the word was presented at study, they were to circle "new."

In addition, for each word that they had recognized as old, subjects were instructed to make a "remember"/"know" judgment (Tulving, 1985; see also Gardiner, 1988; Rajaram, 1993). Essentially, subjects were to label items as "remembered" when they could recollect details of the word's original presentation in the study list (i.e., they could recall perceptual characteristics of the presentation episode, or they remembered having a specific thought when they had encountered the word). "Know" labels were to be assigned when the item failed to bring to mind anything specific about its prior presentation, but the subject nevertheless believed that the item had been presented. The recognition test was self-paced, and subjects were asked to be as accurate as possible.

After completing the recognition test, subjects turned over their response sheets and were given instructions for a modality-judgment test. They were told that they would take another recognition test in which their task was to indicate the presentation modality of the "old" items. They were then given new recognition sheets with the same 144 items that were on the previous recognition test. Next to each item were the letters *NP* (not presented), *P-AUD* (presented auditorily), *P-VIS* (presented visually), and *P-DK* (presented, but I don't know the modality). Subjects were instructed to go through the list as before,

indicating which items they thought were studied and which they thought had not been studied (by circling *NP*). For each studied item, they indicated the prior presentation modality of the item by circling either *P-AUD* or *P-VIS*, or circled *P-DK* if they thought the item had been presented but could not remember the modality. At the end of the session subjects were debriefed and thanked for their participation.

## Results and Discussion

*Recall.* Unless specified otherwise, all differences noted in this article are significant at the .05 level (two-tailed). Recall data are presented in Figure 1, where the probabilities of recalling list items and critical items are presented as a function of serial position and study modality. From Figure 1 it can be seen that recall of list items followed the typical bowed serial position function, with marked primacy and recency effects (e.g., Deese & Kaufman, 1957; Murdock, 1962). Furthermore, items presented auditorily showed a larger recency effect than those presented visually, replicating the traditional modality effect in free recall (e.g., Craik, 1969; Murdock & Walker, 1969; for a discussion, see Crowder, 1976).

A 2 (modality)  $\times$  15 (serial position) analysis of variance (ANOVA) showed that there was no main effect of modality,  $F(1, 51) = 1.71$ ,  $MSE = 0.04$ ,  $p > .10$ , indicating that recall of the list items, collapsing across all 15 serial positions, did not differ between auditory presentation ( $M = .60$ ) and visual presentation (.58). However, there was a significant effect of serial position,  $F(1, 51) = 45.81$ ,  $MSE = 0.04$ , as well as an interaction between serial position and modality,  $F(1, 51) = 3.77$ ,  $MSE = 0.034$ . A paired-observation *t* test confirmed that recall of items from Serial Positions 12–15, the recency portion of the curve, was greater for auditorily presented lists ( $M = .73$ ) than visually presented lists (.62),  $t(51) = 6.58$ ,  $SEM = .017$ . There was also a suggestion that recall of items from primacy and pre-recency serial positions (1–11) was greater for visually presented lists ( $M = .57$ ) than for auditorily presented lists (.55), although this difference did not reach conventional levels of significance,  $t(51) = 1.88$ ,  $SEM = .012$ ,  $p = .07$ . This effect is sometimes obtained but is rarely as large as the effect of modality on recency items (e.g., Craik, 1969; Hintzman, Block, & Inskip, 1972; Murdock, 1966; Murdock & Walker, 1969).

Study modality also affected false recall, as false recall of critical items was significantly greater following auditory presentation ( $M = .46$ ) than visual presentation (.38),  $t(51) = 2.16$ ,  $SEM = .034$ . These results are consistent with those of Smith and Hunt (1998), in that false recall was greater following auditory presentation than visual presentation. The mean number of noncritical intrusions was .39, and the majority of these items were semantically related to at least one item in the list. Interestingly, these noncritical intrusions followed the same pattern as critical intrusions, as noncritical words were intruded more often follow-

<sup>2</sup> Because of a clerical error the critical item *sweet* appeared on the recognition test twice, mistakenly replacing the studied item from Serial Position 8 of that list. Accordingly, we only considered responses to the first appearance of this critical item on the test.

ing auditory presentation ( $M = .45$ ) than visual presentation ( $.32$ ),  $t(51) = 2.44$ ,  $SEM = .051$ .<sup>3</sup>

**Recognition.** Recognition results are reported in Table 1, where the probability of recognizing list items and critical items is presented as a function of study modality of the list and prior task (whether the list was recalled or followed by math problems). These data demonstrate extremely robust levels of false recognition for both modalities, as false alarms to critical items were at least as high as the hit rate in all conditions, whereas false alarms to unrelated lures were rare ( $M = .07$ ). Furthermore, much like list items, critical items were generally judged to be "remembered" more often than "known." Study modality did not appear to affect recognition of list items, and had only a small effect in the predicted direction (auditory > visual) on false recognition of critical items.

A 2 (study modality)  $\times$  2 (item type)  $\times$  2 (task) ANOVA revealed that there was a significant effect of item type,  $F(1, 51) = 4.41$ ,  $MSE = 0.074$ , indicating that critical items were recognized as "old" more often than list items ( $M = .62$  vs.  $.62$ , respectively). There was also a main effect of task,  $F(1, 51) = 15.08$ ,  $MSE = 0.04$ , and a significant interaction between task and item type,  $F(1, 51) = 10.02$ ,  $MSE = 0.022$ . Thus, items associated with previously recalled lists were recognized more often than those associated with lists that were followed by math problems, and this testing effect differed for list items and critical items (as discussed below). There was no main effect of study modality, and all other interactions were not significant (all  $ps > .05$ ). Nonetheless, because we specifically set out to determine the effects of modality, we included this variable in the analyses that follow.

To further explore the interaction between task and item type, we performed separate analyses on critical items and list items. A 2

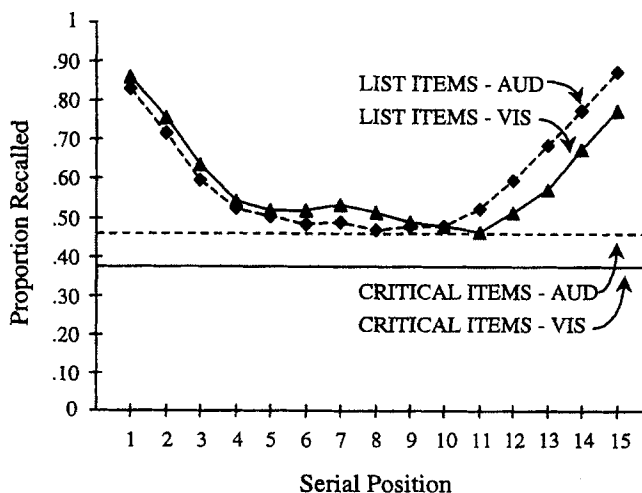


Figure 1. Mean proportion of recalled list items and critical items as a function of serial position and study modality. Recall of list items from auditory lists (AUD) is represented by the diamonds with the dashed curve, and that from the visual lists (VIS) is represented by the triangles with the solid curve. Serial position curves were smoothed by averaging adjacent positions, with the exception of the first (1) and last (15) serial positions, which represent the probability of recall from those positions alone. The horizontal dashed line represents mean false recall of critical items from auditory lists, and the solid line represents that from the visual lists.

Table 1

Average Proportion of List Items and Critical Items Recognized ("Old") in Experiment 1 and Corresponding "Remember" (R) and "Know" (K) Responses, as a Function of Study Modality of the List and Prior Task

Task	Study modality			
	Auditory lists		Visual lists	
	"Old"	(R/K)	"Old"	(R/K)
Study + recall				
List items	.69	(.52/.17)	.67	(.50/.17)
Critical items	.71	(.46/.25)	.68	(.43/.25)
Study + math				
List items	.55	(.34/.22)	.57	(.31/.27)
Critical items	.69	(.40/.29)	.64	(.36/.28)

Note. The false-alarm rate to unrelated lures was .07. Instances where "remember" and "know" judgments do not sum to the proportion of items recognized reflect rounding error.

(study modality)  $\times$  2 (task) ANOVA on critical items revealed a marginally significant effect of study modality,  $F(1, 51) = 3.81$ ,  $MSE = 0.022$ ,  $p = .06$ , indicating that the false-alarm rate to critical items was greater following auditory presentation ( $M = .70$ ) than visual presentation ( $.66$ ). There was no main effect of task and no interaction between task and modality (both  $ps > .10$ ). Although modality affected false recognition in the same direction as that reported by Smith and Hunt (1998, Experiment 2), the effect reported here was quite small. The same analysis on the list items revealed a significant effect of task,  $F(1, 51) = 54.27$ ,  $MSE = 0.014$ , indicating that prior recall boosted recognition of list items relative to the study + math condition ( $M = .68$  vs.  $.56$ , respectively). There was no main effect of study modality and no interaction between modality and task (both  $ps > .10$ ).

The "remember"/"know" results are presented in parentheses in Table 1. As can be seen, subjects were more likely to judge critical items as "remembered" than "known" in all conditions, thereby demonstrating compelling false recollections following presentation in either modality. Furthermore, with reference to unrelated lures, "know" responses ( $M = .06$ ) were greater than "remember" responses ( $M = .02$ ),  $t(51) = 4.26$ ,  $SEM = .009$ , indicating that subjects were not simply responding "remember" by default. We next analyze only "remember" judgments, as these judgments arguably offer a more accurate index of recollective memory than the "old"/"new" judgments.

In general, "remember" judgments followed the same pattern as "old" judgments. The prior task (recall or math problems) considerably influenced "remember" judgments, whereas study modality had small but consistent effects. A 2 (study modality)  $\times$  2 (item

<sup>3</sup> Some of these intrusions may have been due to misperception of items, especially in the auditory condition. We therefore recalculated noncritical intrusions for both auditory and visual lists, excluding any items that did not appear to be related to any list items and instead appeared to have been orthographic or phonological confusions (e.g., *west* for *rest* in the list of items related to *sleep*). Even with this more conservative scoring, intrusions were numerically greater following auditory lists (.36) than visual lists (.29).

type)  $\times$  2 (task) ANOVA on "remember" responses revealed a main effect of modality,  $F(1, 51) = 4.14$ ,  $MSE = 0.025$ , suggesting that subjects were more likely to judge both list items and critical items as "remembered" when the corresponding list had been auditorily presented ( $M = .43$ ) compared with visually presented (.40). There was also a main effect of task,  $F(1, 51) = 42.16$ ,  $MSE = 0.04$ , and an interaction between task and item type,  $F(1, 51) = 14.23$ ,  $MSE = .027$ . Thus, as was the case with the "old" judgments, words were more likely to be judged as "remembered" when their corresponding list had been recalled than if it had been followed by math problems. The interaction suggests that this testing effect was greater for list items ( $M = .51$  vs. .33, respectively),  $t(51) = 10.67$ ,  $SEM = .018$ , than for critical items (.45 vs. .39, respectively),  $t(51) = 2.15$ ,  $SEM = .031$ , collapsing across modality. No other main effects or interactions were significant (all  $ps > .10$ ).

**Modality judgments.** Results from the modality-judgment test are reported in Table 2 and can be easily summarized. First, recollection of study modality was fairly accurate for list items that had been studied either auditorily or visually. This finding is consistent with other studies in the literature that have demonstrated that subjects are quite accurate at recollecting the modality (auditory vs. visual) of studied items even on a surprise test (e.g., Conway & Gathercole, 1987, Experiment 4; Hintzman et al., 1972, Experiment 1; Lehman, 1982). Next, critical items were mistakenly labeled as having been studied in the modality in which their respective lists had been studied, and these errors were made at about the same level as correct judgments of study modality of list items. Finally, prior recall of a list enhanced both accurate and illusory modality judgments. Subjects were more likely to indicate that list items and critical items had been studied in the modality of their study list if that list had initially been followed by recall as opposed to math.

To confirm these observations, paired-observation  $t$  tests indicated that following auditory presentation, both list items and critical items were judged as having been "auditorily" presented

more often than "visually" presented or "don't know." Following visual presentation, both types of items were judged as having been "visually" presented most often (all  $ps < .01$ ). Thus, although subjects were fairly accurate at recollecting the study modality of list items, they also readily assigned the critical item as having been presented in the modality in which its list had been presented. This latter finding was not simply due to subjects' reluctance to respond "don't know" to lures because most of the responses to unrelated lures fell in the "don't know" category ( $M = .06$ ), and this mean was greater than that for both "auditory" (.02) and "visual" (.02) judgments (both  $ps < .001$ ).

In the analyses that follow, to simplify comparisons between list items and critical items and also between task conditions (study + recall or study + math), within each task condition we collapsed modality judgments that were congruent with the actual modality in which the item's list had been presented. Thus, for the list items, we collapsed "auditory" responses to items that had been auditorily studied with "visual" responses to items that had been visually studied, yielding a single measure of modality judgment accuracy for each subject. Similarly, for the critical items we collapsed "auditory" responses for items whose list had been auditorily studied with "visual" responses for those whose list had been visually studied.

A 2 (task)  $\times$  2 (item) ANOVA on these collapsed measures indicated that there was a main effect of task,  $F(1, 51) = 46.92$ ,  $MSE = 0.016$ , and an interaction between task and item type,  $F(1, 51) = 4.54$ ,  $MSE = 0.012$ . Subjects were more likely to judge a list item or critical item as having been presented in the modality of its list when the list had been initially recalled rather than followed by math problems, and this testing effect was greater for list items ( $M = .46$  vs. .31, respectively) than for critical items (.45 vs. .37, respectively). This effect was similar to that found in both the "old"/"new" and "remember"/"know" data from the first recognition test and demonstrates a persistent testing effect. Of more importance, there was no main effect of item type,  $F(1, 51) = 1.92$ ,  $MSE = 0.02$ ,  $p > .10$ , demonstrating that critical items were judged as having been presented in the modality of their lists just as often as list items were. This finding, along with others in the literature, demonstrates that subjects often claim to remember specific details of the critical item's presentation on metamemorial tasks (see also Mather, Henkel, & Johnson, 1997; Payne et al., 1996; Roediger, McDermott, & Pisoni, 2000).

Table 2  
Modality Judgments From Experiment 1, as a Function of Task,  
Study Modality of the List, and Item Type

Task and study modality	Judgments (in proportions)		
	Auditory	Visual	Don't know
Study + recall			
Auditory lists			
List items	.46	.11	.15
Critical items	.48	.13	.13
Visual lists			
List items	.10	.45	.15
Critical items	.11	.43	.18
Study + math			
Auditory lists			
List items	.30	.07	.20
Critical items	.39	.10	.18
Visual lists			
List items	.10	.31	.17
Critical items	.12	.34	.19

Note. The average proportion of unrelated lures judged as having been presented auditorily was .02, visually was .02, and "don't know" was .06.

## Experiment 2

In Experiment 1 we found an effect of study modality on false recall, such that auditory presentation led to more critical intrusions than visual presentation, consistent with the findings of Smith and Hunt (1998) and others (Kellogg, in press; Shobe & Kihlstrom, 1999; Smith, 2000). We found a similar modality effect in false recognition, which is also consistent with the findings of Smith and Hunt (Experiment 2), and suggests that the findings of Maylor and Mo (1999) may have been unique to the particular conditions of their experiment. However, the modality effects on our recognition test were not nearly as large as those reported by Smith and Hunt. Several procedural differences between the two experiments may underlie this difference. Smith and Hunt's subjects studied and recalled six lists of 12 items each before taking the recognition test, whereas our subjects studied and recalled (or

did math after) 24 lists of 15 items each. Further, Smith and Hunt manipulated modality between-subjects, whereas we manipulated modality within-subjects (and in such a way that subjects could not anticipate the modality of an upcoming list). Because our study-recall phase was longer and more complicated than that of Smith and Hunt, our subjects may have been less likely to rely on modality information on the recognition test. To offer a more sensitive test of the modality effects on false recognition, in Experiment 2 we enhanced the salience of study modality by (a) eliminating the recall and math problems following the presentation of each list and (b) blocking lists by study modality (which was again manipulated within-subjects).

A second goal of Experiment 2 was to see if the modality effects found with visual testing could also be obtained with auditory testing. To this end, we manipulated the modality in which items were presented at test as a within-subjects variable. This design allowed for the concurrent manipulation of study and test modality, which could provide additional information about the underlying mechanisms driving modality effects in false recognition. If the study modality effect on false recognition is caused by mechanisms operating solely during the encoding of list items, such as differential activation of critical items, then it should be found regardless of the test modality. However, if the effect involves discrimination processes occurring at retrieval, then test modality might moderate study modality effects. Unlike Experiment 1, all items in Experiment 2 were presented either on the computer screen or over headphones, which allowed for a better match between study and test presentations. This design also allowed us to test whether matching study and test presentation would enhance veridical recognition, as predicted by the encoding specificity principle (e.g., Tulving & Thompson, 1973). Although matching study and test modalities (auditory vs. visual) has not always been shown to enhance recognition on a delayed test (e.g., Craik, Moscovitch, & McDowd, 1994, Experiment 1; Hayman & Rickards, 1995; Hintzman et al., 1972, Experiment 1), at least one study has found recognition to be greater when study and test modalities were matched relative to mismatched conditions (Dean, Yekovich, & Gray, 1988).

## Method

**Subjects.** Thirty-six Washington University undergraduates participated for \$5 each. Data from one subject were replaced due to a failure to follow instructions.

**Materials.** In this second experiment we expanded the set of materials to include all 36 lists found in Stadler et al. (1999). Study lists were constructed in a fashion identical to Experiment 1. Instructions, study stimuli, and test stimuli were all presented through IBM-compatible computers. Study and test words were presented either visually, in the center of the computer screen, or auditorily over headphones. Visual words were presented in lowercase, and auditory words were digitally recorded in a male voice.

Study lists were divided into six sets of six lists each, which were then rotated through the counterbalancing conditions (as discussed below). Using normative data from Stadler et al. (1999), we designed these sets so that (when averaged across the six lists in the set) the average probability of eliciting false recognition of the critical item was equivalent across sets. The average probability of false recognition across the six sets was .66 ( $SD = .01$ ). This assignment method was adopted to reduce the variability across counterbalancing conditions.

**Design.** Both study and test modalities were manipulated within-subjects, resulting in a 2 (study modality: auditory or visual)  $\times$  2 (test modality: auditory or visual)  $\times$  2 (item type: list item or critical item) design. Each subject studied 24 of the 36 lists (four sets), and was then given a recognition test covering all 36 lists. Presentation modality of study lists was blocked, so that half the subjects studied 12 auditory lists and then 12 visual lists, and the other half studied 12 visual lists followed by 12 auditory lists. Items from the 12 nonstudied lists served as the distractors on the recognition test. Subjects were randomly assigned to one of six counterbalancing conditions to control for the possibility of list study-order effects and to ensure that each list was studied auditorily, visually, and nonstudied an equal number of times.

The recognition test consisted of 144 items. Half of these items had been presented during the study phase and half had not been presented. As in Roediger and McDermott's (1995) study, test items were sampled from Serial Positions 1, 8, and 10 from each of the 36 lists, in addition to the 36 critical lures. There were 72 studied items, 3 taken from each of the 24 lists that a subject had studied. The 72 distractors were the three items sampled from each of the 12 lists a subject had not studied (a total of 36 items), plus the 36 critical items from each of the lists (studied or nonstudied). Test items were presented in the same random order for each subject, with the provision that words from any given list would be at least six items apart on the test.

At test an equal number of items was presented auditorily and visually. For items from study lists (and their respective critical lures), half of the items were presented in the same modality as they had been studied and half were presented in the other modality. Similarly, for items from nonstudied lists (and their respective critical lures), half of the test items were presented visually and half were presented auditorily. For counterbalancing, two recognition tests were created; these tests differed only in the modality of the test items, so that all items presented visually on one test were presented auditorily on the other, and vice versa. Half of the subjects in each of the six study conditions were given one recognition test, and half were given the other. Because the modality of test items was randomly ordered, subjects could not anticipate the modality of an upcoming test item.

**Procedure.** Subjects were tested individually by computer and wore headphones during both study and test. They were instructed that they would see and hear 12 lists of 15 words each for a recognition test that would follow, and that they should concentrate during study so as not to miss any of the words. Study lists were blocked, and each list was preceded by a "next list" visual prompt. Additionally, before each set of 12 lists to be studied in the auditory or visual modalities, there was a visual prompt stating "press ENTER to study the 12 auditory lists (or visual lists)." Study items were presented at a rate of approximately 1 s (stimulus onset asynchrony). To equate the exposure duration of items across modalities, visual items were presented on the screen for 500 ms, followed by a blank screen for the remaining 500 ms. Similarly, 500 ms of silence were spliced between each word that was auditorily presented. At test, visual test items were again presented on the screen for 500 ms, and the same digital recordings that were used in the study phase were reused for auditory test items.

Following the presentation of the last list, subjects were given instructions for the recognition test. All subjects had response sheets with the numbers 1–144 arranged in columns to correspond to the test items. Before each test item was presented, a number prompt appeared on the screen to indicate the item's number on the response sheet (i.e., items 1–144). No prompt was given to indicate whether the following item would be presented visually or auditorily. Next to each number on the response sheet were the words "old" and "new," as well as the letters "R," "K," and "G" (for the "remember"/"know"/"guess" judgment). As in Experiment 1, subjects were told that the test consisted of some words they had studied (old) and some words they had not (new). They were further told that some of these words would be presented auditorily and some would be presented

visually and that the presentation modality of a word at test was randomly determined. Regardless of the modality of the item at test, they were instructed to circle "old" if they were confident that they had seen or heard the word at study and "new" if they did not recognize the word from study.

In addition, for each word that they had recognized as "old" subjects were instructed to make a "remember"/"know"/"guess" judgment. Instructions for this judgment were similar to those of Experiment 1, with the addition of the "guess" option (see Gardiner & Conway, 1999). Essentially, "guess" judgments were to be made when subjects recognized a word but could not remember details of its original occurrence and did not have a feeling of knowing that the word had been presented. We added this option for those instances in which a subject judged an item to be "old" based on processes that did not represent the phenomenal states of remembering or knowing, such as feelings of familiarity or fluency, thereby increasing the construct validity of "remember" and "know" responses. When subjects had finished the "old"/"new" and "R"/"K"/"G" judgments for an item they pressed the "enter" key for the next item. Subjects were given as much time as they wished to respond to each item. They were asked to be as accurate as possible and to try to minimize guessing. After completing the recognition test, subjects were debriefed, paid, and thanked for their participation.

### Results and Discussion

The "old"/"new" recognition data are presented in Table 3, with list item data in the top half and nonpresented critical item data in the bottom half. The first point to take from Table 3 is that, as in Experiment 1, we found high levels of false recognition following presentation in either modality. In all conditions, the false-alarm rate to critical items from studied lists ( $M = .69$ , collapsing across modality conditions) approached the hit rate to list items (.71), and these false alarms were considerably greater than those to critical items from nonstudied lists (.35),  $t(35) = 8.67$ ,  $SEM = .038$ . Considering the false alarms to list items and critical items from

nonstudied lists, a 2 (test modality)  $\times$  2 (item type) ANOVA indicated that there was a main effect of item type,  $F(1, 35) = 14.33$ ,  $MSE = 0.024$ , with no effect of test modality and no interaction (both  $ps > .10$ ). The main effect of item type indicates that false alarms to critical items from nonstudied lists ( $M = .35$ , collapsing across test modalities) exceeded false alarms to list items from these same lists (.25), thereby replicating findings by Roediger and McDermott (1995) and Seamon et al. (1998), among others.

In light of these base-rate differences, we used a high-threshold correction procedure to make the comparison between veridical recognition and false recognition more interpretable (e.g., Seamon et al., 1998; Schacter, Verfaellie, & Pradere, 1996). For each subject, the hit rate was corrected by subtracting out false alarms to list items from nonstudied lists, and the false recognition rate was corrected by subtracting out false alarms to critical items from nonstudied lists. This was done separately for items tested auditorily and visually, and the mean corrected scores for each condition are shown in Table 3. Because there were no modality effects on false alarms to list items and critical items from nonstudied lists, both uncorrected and corrected recognition data yielded similar patterns of modality effects.<sup>4</sup>

In general, the effects of study modality on veridical and false recognition were mediated by the modality in which an item was presented at test. When critical items were presented visually at test, auditory study led to greater false recognition than visual study, replicating results by Smith and Hunt (1998) and the small effect found in Experiment 1 (both of which used visual recognition tests). However, when these items were presented auditorily at test, there was no effect of study modality. For list items that were presented visually at test, veridical recognition was greater following visual study than following auditory study. Although Smith and Hunt did not find any modality effects in veridical recognition, there is some evidence in the literature to suggest that matching study and test modalities can enhance recognition (e.g., Dean et al., 1988), and this may have driven the effect of modality on veridical recognition reported here. This modality effect was small, though, and apparently did not affect list items that were tested auditorily.

A 2 (study modality)  $\times$  2 (test modality)  $\times$  2 (item type) ANOVA on corrected recognition scores was consistent with these observations. There was a main effect of item type,  $F(1, 35) = 9.78$ ,  $MSE = 0.105$ , indicating that overall corrected recognition of list items ( $M = .45$ , collapsing across conditions) was greater than that of critical items (.34). The interaction between item type and study modality was also significant,  $F(1,$

Table 3  
Average Proportion of List Items and Critical Items Recognized ("Old") in Experiment 2 and Corresponding "Remember" (R), "Know" (K), and "Guess" (G) Responses, as a Function of Study Modality of the List and Test Modality of the Item

Item type	Auditory test		Visual test	
	"Old"	(R/K/G)	"Old"	(R/K/G)
List items				
Auditory lists	.72	(.43/.17/.11)	.66	(.37/.17/.13)
Visual lists	.72	(.45/.14/.12)	.73	(.45/.15/.13)
Nonstudied lists	.26	(.05/.10/.12)	.25	(.05/.09/.10)
H-T correction				
Auditory lists	.45		.42	
Visual lists	.46		.48	
Critical items				
Auditory lists	.69	(.34/.22/.13)	.75	(.37/.24/.14)
Visual lists	.67	(.31/.19/.18)	.63	(.35/.16/.13)
Nonstudied lists	.36	(.08/.15/.13)	.34	(.06/.08/.20)
H-T correction				
Auditory lists	.33		.41	
Visual lists	.31		.29	

Note. Instances where "remember," "know," and "guess" judgments do not sum to the proportion of recognized items reflect rounding error. H-T correction = data corrected for baseline using the high-threshold procedure (see text).

<sup>4</sup> As an alternative to the high-threshold correction procedure, for each subject we also corrected the hit rate and critical false-alarm rate in each of the relevant modality conditions by calculating  $A'$  (a signal detection estimate of sensitivity based on hits and false alarms, see Donaldson, 1992; Snodgrass & Corwin, 1988). Following Schacter, Israel, and Racine (1999), we first transformed the data in order to avoid hit rates of zero or one, that is,  $p(x)$  was calculated as  $(x + .5)/(n + 1)$ . In those instances in which subjects' hit rates were less than their false-alarm rates, we used the special  $A'$  formula provided by Aaronson and Watts (1987). Analyses of these  $A'$  scores yielded the same pattern of results that was obtained using the high-threshold corrected data in both Experiments 2 and 3. Because these analyses did not add any additional information, we report only the high-threshold corrected data for each experiment.



35) = 10.33,  $MSE = 0.019$ , as was the three-way interaction,  $F(1, 35) = 8.39$ ,  $MSE = 0.013$ . No other main effects or interactions were significant (all  $ps > .10$ ). To further explore these interactions, we first analyzed critical items and list items separately, and then considered the relation between the two.

A 2 (study modality)  $\times$  2 (test modality) ANOVA on corrected false recognition of critical items indicated that there was a significant effect of study modality,  $F(1, 35) = 5.00$ ,  $MSE = 0.035$ , as false recognition was greater following auditory study ( $M = .37$ ) than visual study (.30). There was no main effect of test modality,  $F(1, 35) < 1$ , and the interaction between study and test modality failed to reach significance,  $F(1, 35) = 2.42$ ,  $MSE = 0.032$ ,  $p > .10$ . Although the interaction was not significant, the data suggest that the effect of study modality was greater when critical items were presented visually at test. In this condition, the false-alarm rate to critical items following auditory study ( $M = .41$ ) was significantly greater than that following visual study (.29),  $t(35) = 3.00$ ,  $SEM = 0.039$ , but the smaller effect when critical items were auditorily tested (.33 vs. .31, respectively) was not significant,  $t(35) < 1$ .

With reference to corrected recognition of list items, there were no effects of study modality or test modality (both  $ps > .10$ ), but there was a significant interaction between the two,  $F(1, 35) = 4.34$ ,  $MSE = 0.009$ . As was the case with false recognition, study modality only had a detectable impact on veridical recognition when items were tested visually. With visual presentation at test, list items were recognized more often following visual study ( $M = .48$ ) than auditory study (.42),  $t(35) = 2.59$ ,  $SEM = .026$ , but when tested auditorily there was no effect of study modality,  $t(35) < 1$ .

Another way to assess the effects of modality is to compare levels of corrected false recognition with corrected veridical recognition in each of the four conditions. Looking first at items that were tested visually, in the auditory study condition the recognition of list items ( $M = .42$ ) was not statistically different from false recognition of critical items (.41),  $t(35) < 1$ . In contrast, veridical recognition (.48) exceeded false recognition (.29) in the visual study condition,  $t(35) = 3.63$ ,  $SEM = .053$ . When items were tested auditorily, veridical recognition exceeded false recognition following both auditory study (.45 vs. .33, respectively),  $t(35) = 2.54$ ,  $SEM = .049$ , and visual study (.46 vs. .31, respectively),  $t(35) = 3.08$ ,  $SEM = .049$ . Thus, corrected veridical recognition exceeded corrected false recognition in all conditions except the auditory-study-visual-test condition. In this condition, corrected false recognition was as high as corrected veridical recognition.

The "remember"/"know"/"guess" responses can also be found in Table 3. As seen in Table 3, list items and critical items were predominantly judged "remember" more often than "know" or "guess," again demonstrating compelling false remembering in each modality condition. In contrast, the majority of responses to lures from nonstudied lists fell into the "know" and "guess" categories. Consistent with the "old"/"new" data, a 2 (test modality)  $\times$  2 (item type) ANOVA on "remember" judgments given to lures from nonstudied lists revealed a main effect of item type, as "remember" judgments to critical items from nonstudied lists (.07) were statistically greater than those to list items from these lists (.05),  $F(1, 35) = 4.67$ ,  $MSE = 0.023$ . There was no main effect of test modality and no interaction (both  $ps > .10$ ).

In the same way the "old"/"new" data were corrected, we corrected "remember" judgments given to list items and critical items by subtracting out the "remember" judgments given to their respective control items from nonstudied lists. For simplicity, we analyzed these corrected "remember" responses by comparing those given to critical items with those given to list items in each modality condition. Considering items tested visually, corrected "remember" judgments given to list items ( $M = .32$ ) were not different from those given to critical items (.31) following auditory study,  $t(35) < 1$ , whereas the former (.40) exceeded the latter (.28) following visual study,  $t(35) = 2.80$ ,  $SEM = .041$ . When items were tested auditorily, corrected "remember" judgments were greater for list items than for critical items following both auditory study (.38 vs. .26),  $t(35) = 2.85$ ,  $SEM = .042$ , and visual study (.40 vs. .23),  $t(35) = 3.88$ ,  $SEM = .045$ . Analysis of uncorrected "remember" judgments yielded this same pattern of results. Much like the "old"/"new" recognition data, these results indicate that "remember" judgments given to list items exceeded those given to critical items in every modality condition except for the auditory-study-visual-test condition. Here false recognition was apparently the most compelling, as critical items were judged as "remembered" as often as correctly recognized list items.

### Experiment 3

In Experiment 2 we found an effect of study modality on false recognition, and this effect was predominantly for items that were tested visually. For these items, false recognition was greater following auditory than visual study, consistent with Experiment 1 and Smith and Hunt (1998) but not with Maylor and Mo (1999). Our finding of no modality effects when critical items were auditorily tested was also inconsistent with findings by Maylor and Mo, who found large modality effects on false recognition when items were tested auditorily. As discussed previously, the effects found by Maylor and Mo may have been due to conditions that were unique to their design. Nevertheless, to be confident in our findings in light of this discrepancy, we set out to replicate them in Experiment 3.

Another motivation for Experiment 3 was to examine two hypotheses for modality effects in false recognition: Are they due solely to the access of list-specific modality information at test, or can they also be due to a global distinctiveness heuristic? To test between these accounts, we manipulated study and test modality between-subjects, leaving all other aspects of the design the same as in Experiment 2. Schacter and colleagues found that studying pictures reduced false recognition relative to studying words when presentation format was manipulated between-subjects (Israel & Schacter, 1997) but not when it was manipulated within-subjects (Schacter et al., 1999). They argued that the between-subject manipulation made it possible for subjects to use a distinctiveness heuristic to reduce false recognition. For these subjects, items were studied only in the format particular to their condition. Thus, in the picture condition, if a test item did not bring to mind this distinct pictorial information, it would suggest the item was not studied. In contrast, when study format was manipulated within-subjects, subjects could not be as certain of the format in which an item was presented. As a result, Schacter et al. (1999) argued that these subjects did not base their recognition decisions on the success or failure to recall distinct pictorial information about a test item.

The fact that we manipulated modality within-subjects in Experiments 1 and 2 suggests that our modality effects on false recognition were due to access to list-specific information, as opposed to a global distinctiveness heuristic. However, if modality can also affect false recognition through a global distinctiveness heuristic, then switching from a within-subjects design to a between-subject design should enhance the effect. This is because a between-subject design would allow subjects to use both a distinctiveness heuristic and access to list-specific modality information to reduce false recognition. However, if modality effects on false recognition are due solely to the relative discriminability between memories for items from a specific list and their corresponding critical item, then manipulating study modality between versus within-subjects should have little difference.

### Method

**Subjects.** Eighty-four Washington University undergraduates participated in fulfillment of a course requirement. Subjects were screened to verify they had not participated in any related research.

**Design.** Both study and test modality were manipulated between-subjects, resulting in four conditions with 21 subjects in each: auditory-study-auditory-test, auditory-study-visual-test, visual-study-auditory-test, and visual-study-visual-test. All materials and procedures were identical to those of Experiment 2, with the exception that for each subject, items were studied in a single modality and tested in a single modality, and the instructions were modified accordingly (see below). As in Experiment 2, each subject studied 24 of the 36 lists, and was then given a recognition test covering all 36 lists. Again, items from the 12 nonstudied lists were distractors on the recognition test. Within each of the four modality conditions, subjects were randomly assigned to one of three counterbalancing conditions to control for the possibility of list study-order effects. These conditions were designed so that across subjects, each set of 12 study lists was studied an equal number of times at the beginning or end of the study phase.

**Procedure.** Subjects were tested individually by computer and wore headphones only during the auditory study and test phases. Depending on the condition, they were instructed that they would see or hear 24 lists of 15 words each for a recognition test that would follow. Each list was preceded by a "next list" prompt and there were no other prompts during the presentation of the study lists. All other study instructions and procedures were identical to those in Experiment 2. At test, subjects were told that the words would be presented in the same modality as they had been presented at study or in a different modality, depending on the condition. Subjects were instructed to circle "old" if they recognized the word as one they had studied, regardless of the modality of presentation. The recognition test and all other instructions were the same as those used in Experiment 2.

### Results and Discussion

The results from the between-subjects design are presented in Table 4, with list item data in the top half and nonpresented critical item data in the bottom half. As can be seen, the pattern of false alarms to critical items in this experiment replicated that found using the within-subjects design: Auditory study led to greater false recognition than visual study ( $M = .70$  and  $.62$ , respectively, collapsing across test modality). Veridical recognition also appeared to be affected by modality, as the hit rate was higher in the two conditions where study and test modalities matched ( $M = .72$ ) than in the two conditions where study and test modalities did not match ( $.66$ ). We again found high levels of false recognition relative to veridical recognition in each modality condition, al-

Table 4

*Average Proportion of List Items and Critical Items Recognized ("Old") in Experiment 3 and Corresponding "Remember" (R), "Know" (K), and "Guess" (G) Responses, as a Function of Study Modality of the List and Test Modality of the Item*

Item type	Auditory test		Visual test	
	"Old"	(R/K/G)	"Old"	(R/K/G)
List items				
Auditory lists	.73	(.36/.22/.14)	.69	(.35/.20/.14)
Nonstudied lists	.29	(.04/.10/.15)	.31	(.07/.10/.14)
Visual lists	.63	(.28/.19/.16)	.71	(.39/.17/.15)
Nonstudied lists	.26	(.04/.08/.14)	.24	(.04/.08/.13)
H-T correction				
Auditory lists	.44		.38	
Visual lists	.37		.47	
Critical items				
Auditory lists	.66	(.25/.25/.16)	.73	(.33/.25/.16)
Nonstudied lists	.41	(.04/.13/.24)	.39	(.06/.17/.15)
Visual lists	.59	(.18/.20/.21)	.65	(.23/.24/.18)
Nonstudied lists	.35	(.06/.11/.17)	.42	(.07/.13/.23)
H-T correction				
Auditory lists	.25		.35	
Visual lists	.24		.23	

*Note.* Instances where "remember," "know," and "guess" judgments do not sum to the proportion of recognized items reflect rounding error. H-T correction = data corrected for baseline using the high-threshold procedure (see text).

though as was found in Experiment 2, false alarms to critical items from nonstudied lists ( $M = .39$ ) were more frequent than false alarms to list items from nonstudied lists (.27), collapsing across study and test modality. A 2 (study modality)  $\times$  2 (test modality)  $\times$  2 (item type) mixed-factors ANOVA on these base-rates confirmed that there was a main effect of item type,  $F(1, 80) = 54.90$ ,  $MSE = 0.01$ , but no effects of study or test modality and no two-way interactions (all  $ps > .05$ ). The three-way interaction was marginally significant,  $F(1, 80) = 3.75$ ,  $MSE = 0.01$ ,  $p = .06$ , indicating variability in these base-rates among groups. As a result, we corrected the recognition data using the same high-threshold correction procedure as in Experiment 2.

A 2 (study modality)  $\times$  2 (test modality)  $\times$  2 (item type) mixed-factors ANOVA on corrected recognition scores revealed a significant effect of item type,  $F(1, 80) = 37.08$ ,  $MSE = 0.025$ . This effect indicates that, as in Experiment 2, corrected veridical recognition ( $M = .42$ , collapsing across modality conditions) was greater than corrected false recognition of critical items (.27). There were no other significant main effects or two-way interactions (all  $ps > .10$ ), but the three-way interaction was significant,  $F(1, 80) = 6.44$ ,  $MSE = 0.025$ . As in Experiment 2, we first analyzed modality effects on false recognition and veridical recognition separately and then directly compared false recognition to veridical recognition in each of the conditions.

A 2 (study modality)  $\times$  2 (test modality) ANOVA on corrected false alarms to critical items yielded no main effects or interactions (all  $ps > .10$ ), although this null result may have been due to the inherently weaker between-subject design of Experiment 3. Indeed, inspection of the corrected false recognition data in Table 4 suggests that, as in Experiment 2, study modality did not affect false recognition when items were tested auditorily ( $M = .25$

following auditory study and .24 following visual study) but did appear to have an effect in the predicted direction when critical items were tested visually ( $M = .35$  following auditory study and .23 following visual study). Consistent with these observations, the modality effect on false recognition for items that were tested visually was significant with a one-tailed test,  $t(21) = 1.98$ ,  $SEM = .059$ .

A similar ANOVA on corrected recognition of list items yielded no main effects of study modality or test modality (both  $ps > .10$ ), but the two-way interaction was significant,  $F(1, 80) = 3.85$ ,  $MSE = 0.027$ . This interaction indicates that corrected recognition was greater in the two conditions where test modality matched study modality ( $M = .46$ ) than in the two conditions where there was a modality change from study to test (.38),  $t(82) = 1.98$ ,  $SEM = .036$  (two-tailed). This finding is partially consistent with that of Experiment 2, where it was found that a match between study and test modality yielded greater recognition when list items were tested visually, although a similar effect was not observed for auditorily tested items in that experiment. The discrepancy between Experiments 2 and 3 is not too surprising, as a modality match between study and test has only sometimes been found to facilitate delayed recognition in other list-learning paradigms (as previously discussed).

To directly compare veridical recognition and false recognition, we computed paired-observation  $t$  tests on corrected recognition scores in each of the four conditions. As in Experiment 2, veridical recognition exceeded false recognition in three of the four conditions: visual-study-auditory-test,  $t(20) = 3.45$ ,  $SEM = .04$ , visual-study-visual-test,  $t(20) = 4.96$ ,  $SEM = .047$ , and auditory-study-auditory-test,  $t(20) = 2.86$ ,  $SEM = .066$ . In contrast, veridical recognition ( $M = .38$ ) was not statistically different from false recognition (.35) in the auditory-study-visual-test condition,  $t(20) < 1$ . This pattern replicates that of Experiment 2, again demonstrating that the false recognition effect was greatest in the auditory-study-visual-test condition.

Finally, the "remember"/"know"/"guess" data are also included in Table 4. Although the "remember" judgments to critical items were not as high in Experiment 3 as in Experiment 2, they were still roughly equal to or greater than "know" judgments in each modality condition. In contrast, judgments for list items and critical items from nonstudied lists were predominantly "know" or "guess." A 2 (study modality)  $\times$  2 (test modality)  $\times$  2 (item type) ANOVA on "remember" judgments given to these lures revealed an interaction between item type and study modality,  $F(1, 80) = 6.76$ ,  $MSE = 0.002$ . This interaction reflects the fact that "remember" responses to nonstudied list items were greater following auditory study ( $M = .06$ ) than visual study (.04), whereas those to critical items were greater following visual study (.07) than auditory study (.05), collapsing across test modality. No other main effects of interactions were significant (all  $ps > .05$ ). To correct for this discrepancy, we again took the proportion of "remember" responses to list items and critical items from nonstudied lists and subtracted them from the proportion of "remember" responses given to list items and critical items from studied lists, respectively.

As in Experiment 2, we analyzed these corrected "remember" responses by simply comparing "remember" responses given to list items with those given to critical items in each modality condition. For visually tested items, corrected "remember"

judgments given to list items ( $M = .28$ ) were not different from those given to critical items (.26) following auditory study,  $t(20) < 1$ , whereas the former (.35) exceeded the latter (.16) following visual study,  $t(20) = 4.03$ ,  $SEM = .046$ . For auditorily tested items, corrected remember judgments were greater for list items than for critical items following both auditory study (.32 vs. .21, respectively),  $t(20) = 2.99$ ,  $SEM = .036$ , and visual study (.24 vs. .11, respectively),  $t(20) = 4.61$ ,  $SEM = .029$ . Analysis of uncorrected "remember" judgments yielded the same pattern of results. These findings are consistent with those of the corrected "old"/"new" recognition data of this experiment and also with the findings of Experiment 2. False recognition was most compelling in the auditory-study-visual-test condition, as this was the only condition where subjects claimed to "remember" critical items as often as they claimed to "remember" list items.

### General Discussion

To summarize the primary results, we found that false recall was greater following auditory study than visual study in Experiment 1. Although there was only a small effect of study modality on false recognition on a visual test in Experiment 1, auditory study led to significantly more false recognition than visual study on visual tests in Experiments 2 and 3. Manipulating modality within-subjects (Experiment 2) or between-subjects (Experiment 3) did not influence the magnitude of the modality effect on false recognition. However, modality of test presentation did modulate the modality effect: Auditory study led to greater false recognition only when the critical items were tested visually. Finally, we found that subjects in Experiment 1 were just as willing to erroneously recollect the modality of a critical item's presentation at study as they were to correctly identify a list item's modality. This suggests that false recollection in this paradigm is very compelling and seemingly is accompanied by the retrieval of perceptually specific information. The probability of giving "remember" judgments to critical items was about the same as for list items, confirming this point.

Our finding that auditory study led to greater false recall than did visual study is consistent with the modality effects reported by Smith and Hunt (1998) and others (Kellogg, in press; Shobe & Kihlstrom, 1999; Smith, 2000). The fact that we found a modality effect on visual recognition tests is also consistent with Smith and Hunt. Importantly, the magnitude of our modality effect on false recognition was smaller than that reported by Smith and Hunt, as we found high levels of false recognition in all modality conditions. Although procedural differences may underlie this discrepancy, our experiments are not the first to find robust levels of false recognition following visual study (see Arndt & Hirshman, 1998; Robinson & Roediger, 1997; Seamon et al., 1998).

The similar pattern of corrected false recognition in Experiments 2 and 3 indicates that manipulating modality within-subjects or between-subjects, respectively, made no difference. Considering only items tested visually, the magnitude of the modality effect in Experiment 2 (12%) was identical to that found in Experiment 3 (12%). Similarly, for items that were tested auditorily, we obtained little effect of study modality in Experiment 2 (2%) and Experiment 3 (1%). Because of the overlap between the two experiments, we pooled the main results of Experiments 2 and 3 (using the

corrected data) and they can be found in Figure 2. The figure represents approximately 8,640 observations for list items (2,160 per study–test modality condition) and 2,880 observations for critical items (720 per study–test modality condition).

Turning first to veridical recognition, it can be seen that recognition of list items (collapsing across Serial Positions 1, 8, and 10) was slightly greater in those conditions in which study modality matched test modality relative to those in which study and test modality did not match. This outcome is consistent with the encoding specificity principle and is consistent with the results of Dean et al. (1988), among others. Of course, these effects of modality on list item recognition must be interpreted with some caution because they were driven primarily by Experiment 3 and also because we tested memory for only three items per list and it has been shown that study modality can interact with serial position on a delayed recognition test (Engle & Durban, 1977). Unfortunately, the fact that list items were arranged in order of decreasing relatedness to the critical item in our experiments would have clouded an analysis of the effects of modality on test items from different serial positions.

We turn next to why visual presentation led to lower probabilities of false recall than auditory presentation and also led to lower probabilities of false recognition when critical items were presented visually at test, as shown in Figure 2. One explanation of these effects is that subjects were simply less likely to encode thoughts of the critical item (e.g., as an implicit associate response) or theme of the list (e.g., as a gist representation) during visual study than during auditory study and hence were less likely to later falsely recall or falsely recognize the critical item. However, as Smith and Hunt (1998) pointed out, such accounts are unlikely. There is no evidence in the literature to suggest that implicit associations would be sensitive to study modality, and we know of no previous studies suggesting that thematic or semantic processing of list items should differ between the auditory and visual modalities. Furthermore, our finding that the effect of study modality on false recognition was modulated by test modality suggests that retrieval processes played a role.

A more likely account of the modality effects on false remembering appeals to metamemorial processes that subjects used at the time of retrieval. For simplicity, we first consider the modality effects on false recall, and then turn to the more complicated pattern of effects on false recognition. As discussed in the introduction, memory for visually presented information may be easier

to discriminate from thought at the time of recall than is memory for auditorily presented information (Smith & Hunt, 1998). As a result, subjects would be less likely to erroneously attribute thoughts of the critical item to actual presentation if the list were studied visually, compared with auditorily. In other words, subjects would be more likely to correctly monitor the source of activation of the critical item following visual study, leading to a reduction of false recall. An alternative account of the modality effects on false recall could be posed in terms of the distinctiveness heuristic (Israel & Schacter, 1997; Schacter et al., 1999). For a visually presented list, subjects might be less likely to decide that a critical item had been studied, because the absence of a distinct visual recollection for this item would indicate that it was not presented. Although we manipulated study modality within-subjects in Experiment 1, subjects may have used such a distinctiveness heuristic for each visually presented list because recall always occurred immediately after each list.

Our recognition findings are more difficult to reconcile with this distinctiveness heuristic account, though, because we obtained similar modality effects in false recognition with the within-subjects design (Experiment 2) and the between-subjects design (Experiment 3). As discussed previously, subjects should have been able to use a distinctiveness heuristic only when modality was manipulated between-subjects because only these subjects could have been confident that all of the studied items should have brought to mind detailed visual recollections (see Schacter et al., 1999). Instead, our pattern of results is more consistent with the idea that subjects accessed list-specific modality information and used this information to reduce false recognition from visually studied lists. Encountering a critical item at test may have served as a cue for some of the studied items from its list, along with corresponding modality-specific information. The finding in Experiment 1 that subjects were generally accurate at recollecting the presentation modality of list items suggests that such information was accessible to them. Assuming that visual information is more discriminable from thought than is auditory information, subjects would be better at reducing false recognition for visually presented lists than for auditorily presented lists.

The fact that we found modality effects for critical items tested visually but not for those tested auditorily complicates the story. Apparently, the metamemorial processes that subjects used to reduce false recognition were sensitive to both study and test modality. One tentative explanation is that subjects were more

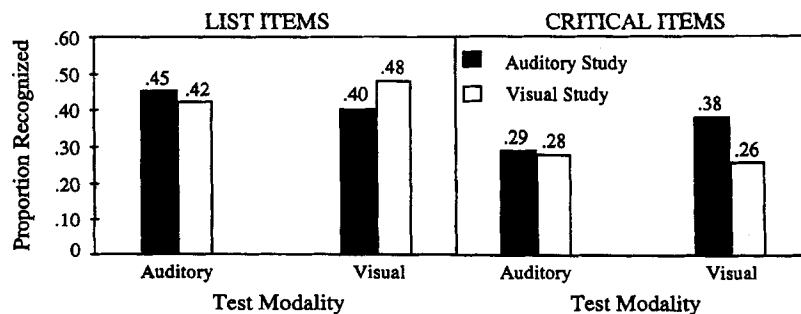


Figure 2. Mean corrected recognition of list items (left) and of critical items (right), averaged across Experiments 2 and 3, as a function of study modality of the list and test modality of the item.

likely to use reality monitoring processes to reduce false recognition from visual lists when critical items were presented visually at test. Visual test presentation provides subjects with the appearance of the critical item as it would have looked if it had been visually presented at study, and this might make the lack of visual recollections for this item more salient. As a result, subjects might have been more likely to use the absence of this information as indicative that the critical item was not presented at study.

There is some evidence in the literature to suggest that source-monitoring processes can be sensitive to the relative salience of modality information at test. Marsh and Hicks (1998, Experiment 2) found that subjects were better at rejecting lures on a source-monitoring task when the question focused attention to visual characteristics from study ("Did you see this item?") as opposed to auditory characteristics ("Did you hear this item?"). From this finding they argued that

the lack of visual details in memory, corroborated by a visual [test] display that is especially visually unfamiliar, leads to better performance when subjects are focused on visual details. One straightforward prediction from this result is that not only should source-monitoring performance covary with the sources presented at study, but that it may also be a complex function of the test modality. (p. 1144)

Our findings are consistent with this prediction, in that our subjects seemed to use the lack of visual details to reject critical lures only when these lures were presented visually at test.

Other findings in the literature suggest that test modality can modulate the effects of study modality on false remembering. Israel and Schacter (1997) found an analogous testing effect in their picture + word condition. In this condition, false recognition was less likely if critical items were presented pictorially at test (i.e., both the picture and its auditory label were presented) than if they were only presented auditorily at test. Their account for this testing effect was similar to ours discussed above: Presenting the critical item pictorially at test made the absence of such a distinct representation more salient than presenting the critical item only verbally, thereby increasing the likelihood that subjects would invoke metamemorial strategies to reduce false recognition.<sup>5</sup>

Data recently presented by Kellogg (in press, Experiment 1) are also relevant. Subjects studied DRM lists for immediate free recall. Presentation modality of each list (auditory vs. visual) was manipulated within-subjects. Importantly, recall modality was also manipulated within-subjects, with some lists followed by written recall and others by spoken recall. Consistent with our results, on the written recall test Kellogg found that false recall was greater following auditory study (.40) than visual study (.21). On the spoken recall test, however, study modality had no effect on false recall (.30 for auditory, .31 for visual). Apparently, the written recall test required subjects to access orthographic codes that are specific to the visual modality, whereas spoken recall did not. As a result, written recall may have made the absence of visually encoded information for the critical items more salient, resulting in a modality effect for written but not spoken recall.

We proposed that test modality may influence false remembering by modulating the retrieval and subsequent use of study modality information in metamemorial decision processes. Although this account is plausible and is consistent with other re-

search, it does not adequately handle all aspects of our data. Specifically, if visual presentation at test helped subjects reduce false recognition of critical items from visual study lists, then false recognition should have been lower in the visual-study-visual-test condition than in the visual-study-auditory-test condition and the auditory-study-auditory-test condition. In fact, false recognition was roughly equivalent across these three conditions in both Experiments 2 and 3. We have no good explanation for this particular pattern, except that perhaps auditory testing leads to less false recognition in this paradigm than visual testing, all other things being equal. This outcome is theoretically challenging, and future research is needed to further develop the present account, or perhaps to provide a different account of the interaction between study and test modality on false recognition. Whatever the explanation, our findings indicate that both encoding and retrieval factors must be considered for a complete understanding of modality effects on false remembering.

We turn lastly to the modality judgments of Experiment 1 (see Table 2), along with their implications. The finding that critical items were assigned to a study modality as often as list items, even though subjects had the option to respond "don't know," is open to several possible accounts. The first is that subjects imagined the critical item's occurrence during the study phase, and this imagination was subjectively experienced as a presentation in the modality of the respective list. We view this explanation as highly unlikely because there is no particular reason to believe that the presentation of these materials results in perceptual imagery of nonpresented words. However, McKone and Murphy (2000) recently demonstrated that visually presenting associative lists led to priming of nonstudied critical items on a visual implicit test, whereas auditory study of the lists did not. To explain this pattern, McKone and Murphy suggested that thoughts of the critical item at study might be specific to the modality of the list, resulting in priming of the critical item only when the test reinstates this modality. As these authors discuss, though, there are other potential explanations for these modality-specific priming effects. Thus, although they are intriguing, further work needs to be done before accepting these priming effects as implicating spontaneous imagery of the critical item at study.

A second, and more plausible, explanation of the present modality judgment findings is that subjects thought of the critical item during the study phase, but its activation was not associated with any modality-specific information (e.g., it was simply worked into the mental rehearsal of the list). In this case, subjects may have determined the study modality for the critical item at test, by accessing the modality of the relevant list and then attributing the critical item to having occurred in this modality. These erroneous attributions might have been especially compelling if, as part of the attribution process, the

<sup>5</sup> It should be noted that Israel and Schacter (1997) did not find a significant effect of test presentation (auditory + visual vs. auditory alone) in their word condition (where items were studied both auditorily and visually), which is arguably more relevant to the conditions of the present experiment. However, if one corrects their data by subtracting the relevant base-rates reported in their Table 1, the overall effect is in the predicted direction.

subject imagined the critical item (at test) as having occurred in the relevant modality. A similar process could have driven these modality judgments even if the critical item was not generated at study, but instead was falsely recognized solely because of its resemblance to the theme of the list, and subsequently was thought to have occurred in the modality of that list. In either case, erroneous modality judgments would be the result of attributing present thoughts and feelings to past occurrences (i.e. a memory attribution; see Jacoby, Kelley, & Dywan, 1989).

A final explanation is that information in memory that supported false recognition was retrieved in combination with perceptual information encoded from presented items. Because of this mixture of retrieved information, subjects may have subjectively experienced illusory recollection of the critical item in a particular modality, in the absence of attributions or imaginative processes. An account such as this has been proposed within fuzzy-trace theory, in that the retrieval of strong gist representations may be associated with vivid perceptual details that are characteristic of verbatim traces (what Brainerd et al., 2001, call "phantom recollection"). The exact mechanism for this type of process has yet to be specified, and as such, it is unclear to us how this proposal advances our knowledge beyond the attributional framework of Jacoby et al. (1989). Nonetheless, as with the previous explanation, the main point to be made here is that these seemingly objective modality judgments appeared to be quite sensitive to reconstructive retrieval processes.

In sum, we replicated Smith and Hunt's (1998) study by demonstrating that false recall (Experiment 1) and false recognition on visual tests (Experiments 2 and 3) were greater following auditory study than visual study. Because the same pattern of modality effects was obtained with a within-subjects and a between-subjects design, the reduction in false recognition following visual presentation appears to be due to access to list-specific modality information rather than to the use of a global distinctiveness heuristic. We also found that the effects of study modality on false recognition were obtained when critical items were presented visually at test but not when they were presented auditorily. Regardless of the specific theory used to account for these modality effects, this finding suggests that theories that appeal solely to processes occurring at study are inadequate.

Finally, we found high levels of false remembering relative to veridical remembering in all conditions, demonstrating that this illusion can be robust following presentation in either modality. We also confirmed that false recognition is often accompanied by specific recollections of illusory details, as indicated by high levels of "remember" judgments and modality judgments for critical items. These latter findings add to the growing body of evidence that memories for events that did not occur can sometimes be judged to have similar subjective detail as memories for those events that did occur. Such findings force us to consider that all explicit judgments about past experiences, including detailed judgments such as "remember"/"know" or modality, as well as less-detailed "old"/"new" judgments, are influenced by reconstructive processes. As a result, they should be quite sensitive to various metamemorial strategies, such as those discussed here in light of the effects of modality on false remembering.

## References

- Aaronson, D., & Watts, B. (1987). Extension of Grier's computational formulas for A' and B' to below-chance performance. *Psychological Bulletin*, *102*, 439-442.
- Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language*, *39*, 371-391.
- Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 307-327.
- Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*, 341-361.
- Craik, F. I. M. (1969). Modality effects in short-term storage. *Journal of Verbal Learning and Verbal Behavior*, *8*, 658-664.
- Craik, F. I. M., Moscovitch, M., & McDowd, J. M. (1994). Contributions of surface and conceptual information to performance on implicit and explicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 864-875.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Dean, R. S., Yekovich, F. R., & Gray, J. W. (1988). The effect of modality on long-term recognition memory. *Contemporary Educational Psychology*, *13*, 102-115.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17-22.
- Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, *54*, 180-187.
- Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275-277.
- Engle, R. W., & Durban, E. D. (1977). Effects of modality of presentation on delayed recognition. *Perceptual & Motor Skills*, *45*, 1203-1210.
- Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, *4*, 271-276.
- Gallo, D. A., & Roediger, H. L., III. (2000). *Variability among word lists in evoking associative memory illusions*. Manuscript in preparation.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309-313.
- Gardiner, J. M., & Conway, M. A. (1999). Levels of awareness and varieties of experience. In B. H. Challis & B. M. Velichovsky (Eds.), *Stratification in cognition and consciousness* (pp. 237-254). Amsterdam: John Benjamin Publishing.
- Hayman, C. A. G., & Rickards, C. (1995). A dissociation in the effects of study modality on tests of implicit and explicit memory. *Memory & Cognition*, *23*, 95-112.
- Hintzman, D. L., Block, R. A., & Inskip, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, *11*, 741-749.
- Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*, 577-581.
- Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67-85.
- Kellogg, R. T. (in press). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Lehman, E. B. (1982). Memory for modality: Evidence for an automatic process. *Memory & Cognition*, *10*, 554-564.
- Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-

- monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1137–1151.
- Mather, M., Henkel, L. A., & Johnson, M. J. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*, 826–837.
- Maylor, E. A., & Mo, A. (1999). Effects of study-test modality on false recognition. *British Journal of Psychology*, *90*, 477–493.
- 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. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language*, *39*, 508–520.
- McDermott, K. B., & Watson, J. M. (in press). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory & Language*.
- 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.
- Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, *36*, 513–520.
- Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488.
- Murdock, B. B., Jr. (1966). Visual and auditory stores in short-term memory. *Quarterly Journal of Experimental Psychology*, *18*, 206–211.
- Murdock, B. B., Jr., & Walker, K. D. (1969). Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 665–676.
- Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*, 261–285.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*, 89–102.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75.
- Rhodes, M. G., & Anastasi, J. S. (2000). The effect of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin & Review*, *7*, 158–162.
- Robinson, K. J., & Roediger, H. L., III. (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*, 231–237.
- Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and the arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association.
- Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.
- Roediger, H. L., III, McDermott, K. B., & Pisoni, D. B. (2000). *Recollection of illusory voices*. Manuscript submitted for publication.
- Roediger, H. L., III, McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187–245). Hove, England: Psychological Press.
- Russell, W. A., & Jenkins, J. J. (1954). *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test* (Tech. Rep. No. 11, Contract N8 ONR 66216, Office of Naval Research). University of Minnesota, Twin Cities Campus.
- Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1–24.
- Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language*, *35*, 319–334.
- Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for non-conscious processes. *Psychological Science*, *9*, 20–26.
- Shobe, K. K., & Kihlstrom, J. F. (1999, November). *The role of list strength, structure, and presentation modality in recollective experience accompanying false memory*. Poster presented at the 39th Annual Meeting of the Psychonomic Society, Los Angeles.
- Smith, R. E. (2000, April). *The effect of age and presentation modality on false recall*. Poster presented at the 8th Biennial Cognitive Aging Conference, Atlanta, GA.
- Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, *5*, 710–715.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.
- Stadler, M. A., Roediger, H. L., III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*, 494–500.
- Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language*, *33*, 203–217.
- Thapar, A., & McDermott, K. B. (in press). False recall and false recognition induced by presentation of associated words: Effects of level of processing and retention interval. *Memory & Cognition*.
- Toglia, M. P., & Neuschatz, J. S. (1996, November). *False memories: Where does encoding opportunity fit into the equation?* Poster presented at the 37th Annual Meeting of the Psychonomic Society, Chicago.
- Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, *7*, 233–256.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, *26*, 1–12.
- Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.
- Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122–129.

Received December 22, 1999  
 Revision received August 24, 2000  
 Accepted August 24, 2000 ■