

The modality effect in false recognition: Evidence for test-based monitoring

BENTON H. PIERCE

Texas A&M University, Commerce, Texas

DAVID A. GALLO

University of Chicago, Chicago, Illinois

and

JONATHAN A. WEISS and DANIEL L. SCHACTER

Harvard University, Cambridge, Massachusetts

False recognition in the Deese/Roediger–McDermott (DRM) paradigm has been shown to be greater following auditory study than following visual study, but there are competing explanations for this effect. We generalized this phenomenon in Experiment 1, finding an equivalent modality effect for associative (DRM) lists and categorized lists. Because conscious generation and subsequent monitoring of related lures during study is infrequent for categorized lists, this result is inconsistent with the idea that the modality effect is due to a study-based monitoring process. An alternative explanation is that visual study impairs relational processing relative to auditory study, which could cause a modality effect by lowering false recognition of related lures. We tested this idea in Experiment 2, by switching to a meaning-based test that is sensitive only to the retrieval of relational information. A modality effect was not obtained for either type of list on this test. The results from both experiments were predicted by a test-based monitoring account, rather than by the study-based monitoring or relational processing accounts.

The study of false memories has assumed increasing empirical and theoretical importance in memory research. Numerous studies have begun to elucidate both the cognitive properties and neural substrates of various types of false memories, and a variety of theoretical accounts have been proposed (for recent reviews, see Gallo, in press; Schacter & Slotnick, 2004). During the past several years, increased emphasis has been placed on whether such memories are attributable to processes that occur during encoding or during retrieval (see, e.g., Dodd, Sheard, & MacLeod, 2006; Marsh, McDermott, & Roediger, 2004; S. M. Smith, Gerken, Pierce, & Choi, 2002). An important aspect of this study-based versus test-based issue is the role that metacognitive monitoring mechanisms play in the reduction of false memories. Here we use the term *monitoring* in a general sense, to refer to the various inference or decision processes that participants use to enhance the accuracy of their memory judgments (see, e.g., Johnson, Hashtroudi, & Lindsay, 1993; Koriati & Goldsmith, 1996; Schacter, Norman, & Koutstaal, 1998). These processes are typically thought to occur during retrieval, but of course, encoding factors play a critical role, by determining what information will be available at retrieval, and

hence inform decision processes. For instance, Schacter et al. have argued that studying distinctive perceptual information can help participants avoid familiarity-based false recognition, by relying more on detailed recollections at test (Dodson & Schacter, 2001, 2002; Israel & Schacter, 1997; Schacter, Cendan, Dodson, & Clifford, 2001; Schacter, Israel, & Racine, 1999).

Another finding in which monitoring processes may play an important role is the phenomenon that false memories can differ as a function of presentation modality (Gallo, McDermott, Percer, & Roediger, 2001; Kellogg, 2001; R. E. Smith & Hunt, 1998). Roediger and McDermott (1995), using a method developed by Deese (1959), presented participants with lists of words (e.g., *candy*, *sour*, *sugar*) that are all semantically related to a non-presented theme or lure word (e.g., *sweet*). On an immediate free recall test, participants often falsely recalled the theme words, with intrusion rates approximately equal to the probability of recalling words from the middle of the list. On a subsequent recognition test, false alarm rates to the critical lures were comparable to the hit rates for studied items. Using the Deese/Roediger–McDermott (DRM) false memory method, Gallo et al. (2001), Kellogg (2001), and R. E. Smith and Hunt (1998) found that false recall and false recognition from DRM lists were lower following visual presentation than following auditory presentation at study, with written recall and visual recognition tests.

To explain the effect of presentation modality on false recall and false recognition, R. E. Smith and Hunt (1998) suggested that visual presentation of study words promotes

This research was supported by National Institute on Aging Grant AG08441. We appreciated the helpful comments of Reed Hunt, Kathleen McDermott, and Elinor McKone. Correspondence concerning this article should be addressed to B. H. Pierce, Department of Psychology and Special Education, Texas A&M University, Commerce, TX 75429 (e-mail: benton_pierce@tamu-commerce.edu).

more distinctive item-specific processing than does auditory processing. As a result, participants are better able to discriminate between visual events (studied words) and internally generated events (the related lures) than they are between auditory events and internally generated events. That is, visual presentation leads to better reality monitoring (Johnson & Raye, 1981) than does auditory presentation. Such monitoring could occur at study, at test, or during both phases. During study, participants may generate the related lure and realize that it was not actually presented, thereby avoiding subsequent false recognition. During test, participants may demand more distinctive recollections of an item before producing it on a recall test or saying "old" to the item on a recognition test. Because nonstudied lures should elicit less distinctive recollections, they would be easier to reject (see, e.g., Gallo & Roediger, 2002).

Consider the study-based monitoring process. This mechanism appeals to processes that operate at encoding, and it is derived from association-based theories of false memory in the DRM paradigm (Gallo, in press; McDermott & Watson, 2001; Roediger, Watson, McDermott, & Gallo, 2001). According to association-based theories, studying a list of semantic associates (e.g., *candy, sour, sugar*) activates the mental representation of the critical nonpresented word *sweet*. That is, *sweet* is an implicit associative response as conceptualized by Underwood (1965). Consequently, false memory occurs when this association-based activation is mistaken for actual presentation of the word, which represents a reality-monitoring error (Johnson & Raye, 1981). Consistent with the findings of R. E. Smith and Hunt (1998), discriminating an activated critical lure representation during the encoding of DRM word lists might be easier when study items are presented visually than when they are presented auditorily.

Gallo et al. (2001) also appealed to monitoring to explain modality effects, but argued that test-based processes are important to consider. Gallo et al. found lower false recall and false recognition following visual presentation than following auditory presentation. However, modality effects in false recognition were found only for critical items tested visually, not for those tested auditorily. Because they found an effect of test modality, Gallo et al. argued that retrieval factors (occurring at test) played a role in the study modality effect. Following R. E. Smith and Hunt (1998), Gallo et al. proposed that memory for visually presented information may be more discriminable from thought than is memory for auditorily presented information. To explain how this discrimination could reduce false recognition, Gallo et al. appealed to the idea of a *distinctiveness heuristic* (Dodson & Schacter, 2001, 2002; Schacter et al., 2001; Schacter et al., 1999). The distinctiveness heuristic is a mode of responding in which participants base their responses on the metamemorial assessments of the kinds of information they feel they should remember. Israel and Schacter (1997) showed that participants who had studied list words in a more distinctive format (pictures, in this case) exhibited lower false recognition than did participants who had studied words. They argued that participants in the distinctive condition

could expect more distinctive recollections at test. Similarly, Gallo et al. argued that visual study of words is more distinctive than auditory study, and could facilitate the use of a distinctive heuristic at test.

Importantly, Gallo et al. (2001) obtained similar reductions in false recognition following visual study in both a between-subjects and a within-subjects design. Schacter et al. (1999) argued that a global distinctiveness heuristic should operate only in a between-subjects design, because only in this instance can participants be confident that all studied items should produce recollections of distinctive visual information. The fact that Gallo et al. found similar modality effects in both designs suggests that participants used a more local, list-specific distinctiveness heuristic. This idea is supported by Gallo et al.'s findings (in Experiment 1) that participants were generally accurate at recollecting the presentation modality of study items, suggesting that this list-specific modality information was accessible. Likewise, Kellogg (2001) utilized a within-subjects design and found that false written recall was significantly lower following visual presentation than after auditory study, suggesting that participants can use distinctive orthographic information provided by visual study to avoid critical intrusions that lack such detail. When participants were instructed to visually imagine auditorily presented words, the modality effect was eliminated. The results of Gallo et al. and Kellogg support the idea that the effects of study modality in false recall and false recognition are driven primarily by better reality monitoring processes, such as a list-specific distinctiveness heuristic, afforded by visual as opposed to auditory presentation.

Although these explanations of Gallo et al. (2001) and Kellogg (2001) can clearly fit the data, other alternatives are viable. For example, Arndt and Reder (2003) have recently offered an alternative account of false recognition modality effects, based on the distinction between item-specific and relational processing (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993). The item-specific and relational processing view holds that relational information in memory stems from processing the similarities among study items, whereas memory for item-specific information is generated from processing differences among study items. According to Arndt and Reder, relational processing boosts recognition of study items, but also leads to increased false recognition of related lures, through the encoding of overlapping semantic features or gist information (see Brainerd, Wright, Reyna, & Payne, 2002). They argued that increased item-specific processing may result in reduced relational processing, due to limited attentional resources. Therefore, conditions that promote the encoding of distinctive, differentiating information, such as visual presentation, should reduce relational processing and hence, reduce false recognition. According to this account, modality effects in false recognition can be explained by relational or gist-based processes, with no need to appeal to decision-based (i.e., monitoring) processes.

In this paper, we describe two experiments designed to test these alternative accounts of the modality effect in

false recognition. In Experiment 1, we pitted study-based monitoring against the test-based monitoring account. This experiment used both associative (i.e., DRM) and categorized word lists. As discussed below, previous research indicates that these two types of materials differ with regard to the likelihood that conscious thoughts of critical, nonpresented lures will be generated during study. Our rationale in using these different materials is that, to the extent that study-based monitoring processes underlie the effect of presentation modality on false recognition, we should see dissociations between list type and the level of the effect. That is, materials that are more likely to elicit thoughts of critical lures during study, and therefore afford more opportunity for study-based monitoring, should result in a greater modality effect. In contrast, if test-based monitoring is the primary mechanism, similar modality effects should be observed. In Experiment 2, we contrasted test-based monitoring with the relational processing account, by switching to a test that should be sensitive to the latter but not the former.

EXPERIMENT 1

Experiment 1 was designed to examine study-based versus test-based monitoring accounts of the modality effect in false recognition. Our rationale involved the utilization of two fundamentally different types of word lists, associative lists and categorized lists, that are likely to vary in their probability of consciously activating critical nonpresented lure items during study. A number of associative lists can elicit activation of critical items, although this activation has been shown to vary substantially among different lists (see, e.g., Deese, 1959; Gallo & Roediger, 2002; McEvoy, Nelson, & Komatsu, 1999). This variability is due, in large part, to differences in mean backward associative strength (MBAS) among word lists (Deese, 1959). BAS is the probability that a list item will evoke the critical item as a response in a free association task. Whereas associative lists with high MBAS are likely to consciously activate the critical lure at study, categorized lists, in general, have low MBAS and are unlikely to strongly activate a critical nonpresented lure at study (S. M. Smith et al., 2002). S. M. Smith et al. found that in a free association task, associative lists tended to evoke critical lures as responses, whereas categorized lists did not. In a word stem completion task, moreover, indirect priming effects for critical items were found for associative lists, but not for categorized lists (see also McDermott, 1997, for indirect priming effects using associative lists). In the present study, we used differences in conscious activation of critical lures between associative and categorized lists to test the study-based monitoring explanation of the modality effect. If the modality effect is due to monitoring critical lures during study, then unlike for associative lists, we should find little or no modality effect for categorized lists, because critical lures are not generated and monitored during study. If the modality effect is due to test-based monitoring, then we should obtain the same effect for both types of lists. In this case, participants

should be able to access list-specific modality information at retrieval and use a list-specific distinctiveness heuristic to correctly reject lures from visually presented lists. Regardless of whether the lists differ in conscious activation of critical lures during study, both types of lists elicit false recognition that could potentially be monitored at test.

Method

Participants. A total of 40 university students were recruited using signs posted on the Harvard University campus or from the Harvard subject pool.

Materials and Design. Twenty-four categorized lists and 24 associative lists of 15 items each were used. Categorized lists were taken from the Battig and Montague (1969) category norms, whereas associative (DRM) lists were taken from Roediger, Watson, et al. (2001). MBAS was significantly greater for the associative lists ($M = .174$) than for the categorized lists ($M = .066$) [$t(23) = 5.88$, $SEM = .018$].

The 48 lists were divided into two sets of 24 for counterbalancing purposes. Participants studied 24 lists (12 associative and 12 categorized) and then completed a recognition test containing items from the 48 lists. Because of variability in associative strength, the two counterbalancing sets of associative lists were matched in terms of average BAS, as were the two sets of categorized lists. As in Roediger and McDermott (1995), words in the associative lists were arranged in descending order of associative strength to the nonstudied critical item. As in Seamon, Luo, Schlegel, Greene, and Goldenberg (2000), categorized list words were arranged in descending order of response frequency according to the Battig and Montague (1969) norms, and the most frequent exemplar from each list was omitted and served as the critical item.

Study modality and list type were manipulated within subjects, resulting in a 2 (modality: auditory or visual) \times 2 (list type: categorized or associative) design. The 24 study lists were presented to each participant in the same random order, with 12 lists presented auditorily and 12 presented visually. Categorized and associative lists were interspersed and presented in an alternating fashion. Presentation modality was blocked, so that half the participants studied 12 auditory lists and then 12 visual lists, and the other half studied 12 visual lists and then 12 auditory lists. Participants were randomly assigned to one of eight counterbalanced conditions so that each list was studied auditorily, studied visually, and nonstudied an equal number of times.

Following presentation of the study lists, participants completed a 192-item visual recognition test. A visual test was used because the visual test modality yields the strongest study modality effects and thus has been theorized to engender the most monitoring (see Gallo et al., 2001; Kellogg, 2001; Roediger, McDermott, Pisoni, & Gallo, 2004). The test contained 72 targets or studied words (from Positions 1, 8, and 10 of the studied lists), 72 target controls or unrelated lures (from Positions 1, 8, and 10 of the nonstudied lists), 24 false targets or related lures (the critical lures from each of the studied lists), and 24 false-target controls or unrelated lures (the critical lures from each of the nonstudied lists). No more than three test items of the same type appeared consecutively, and items were randomly intermixed.

Procedure. Participants were tested individually using a computer. They were instructed that they would first hear (or see) a set of word lists and then see (or hear) a second set of lists. The participants were urged to pay close attention to the words in preparation for a later, unspecified memory test. Study lists were blocked, and each list was preceded by an auditory or visual "next list" prompt. Before each set of 12 lists to be studied in the auditory or visual modalities, there was a visual prompt stating "press the space bar to begin." Study items were presented auditorily (in a female voice) through computer speakers or visually (in lowercase letters) in the center of the computer screen. Items were presented every 3.25 sec

in both conditions. Visually presented words were presented for 3 sec, followed by a 250-msec delay before the next word appeared. After the last list was presented, the participants were given instructions for a recognition memory test. They were instructed to indicate whether an item was “old” (i.e., it had been heard or seen on one of the 24 study lists) or “new” (i.e., it had not been heard or seen on any of the lists) by pressing the keys that had been labeled “O” and “N,” respectively. The participants were told to press the *old* key only for the words that they were reasonably certain had appeared on the lists. All words were presented visually in the middle of the screen, with “Old or New” appearing below the word. The participants were also asked to make *remember/know* judgments for each word they had recognized by pressing keys that had been labeled “R” and “K,” respectively. For the remember/know instructions (see Rajaram, 1993), participants were asked to label an old item as *remembered* if they could consciously relive some aspect of what happened or what was experienced at the time they initially heard or saw the word. That is, a remembered word should bring to mind a particular image or association from the time of study, or the way the word sounded or its position in the lists. The participants were to assign a *know* label to a word if they were certain it had been presented, but could not consciously relive anything about the word’s actual occurrence (such as what it looked or sounded like). Examples of both remember and know judgments were given for further clarification. The recognition test and remember/know judgments were self-paced, and the participants were instructed to be as accurate as possible. After the test, the participants were debriefed and thanked for their participation.

Results and Discussion

Unless noted otherwise, a significance level of $p < .05$ was used on all statistical tests in this study. *Old/new* recognition data are presented in Table 1, with nonstudied critical item data in the top half and list item data in the bottom half. Following Gallo et al. (2001), we corrected the false recognition rate by subtracting false alarms to critical items from nonstudied lists, and the true recognition rate by subtracting false alarms to targets from nonstudied lists. Unless analyses on uncorrected and corrected recognition levels differ, analyses on corrected recognition (both true and false) will be reported throughout the rest of the article.

The first point to note from Table 1 is the between-list difference in false alarm rates to critical items. A 2 (list type) \times 2 (study modality) repeated measures ANOVA on corrected false recognition revealed a main effect of list type [$F(1,39) = 29.16$, $MS_e = .054$]. False alarms to critical items from associative lists ($M = .48$, collapsed across modalities) were significantly greater than to those from categorized lists ($M = .28$). This pattern is consistent with the notion that critical items are activated to a greater extent in associative lists than in categorized lists. The ANOVA also revealed a main effect of modality [$F(1,39) = 10.72$, $MS_e = .053$]. False alarms to critical items were greater when lists were studied auditorily ($M = .44$, collapsed across lists) than when they were studied visually ($M = .32$). More importantly, there was no study modality \times list type interaction [$F(1,39) < 1$], indicating that the size of the study modality effect was equivalent for categorized and associative lists.

A 2 (study modality) \times 2 (list type) ANOVA on corrected true recognition revealed a main effect of list type

Table 1
Experiment 1: Proportions of List Items and
Critical Items Recognized (Old) and Corresponding
Remember (R) and Know (K) Responses, as a Function
of Study Modality of the Lists and List Type

	List Type			
	Categorized		Associative	
	Old	R/K	Old	R/K
Critical Items				
Auditory lists	.44	.20/.80	.65	.48/.52
Visual lists	.32	.26/.74	.54	.34/.66
Nonstudied lists	.10	.13/.87	.11	.20/.80
Corrected				
Auditory lists	.35		.54	
Visual lists	.22		.43	
List Items				
Auditory lists	.71	.56/.44	.65	.62/.38
Visual lists	.72	.64/.36	.70	.67/.33
Nonstudied lists	.05	.14/.86	.09	.15/.85
Corrected				
Auditory lists	.66		.57	
Visual lists	.67		.62	

[$F(1,39) = 8.16$, $MS_e = .025$] indicating that recognition of list items was higher for categorized lists ($M = .66$) than for associative lists ($M = .59$). This is important because it suggests that greater false alarms to related lures from associative lists are not simply due to an increased willingness of participants to say “old” to items from associative lists. There was no main effect of modality or interaction (both $ps > .27$).

Table 1 also shows responses from the remember/know task. Because of list differences in false alarms to critical items, we expressed these responses as proportions of “old” responses. Note that these subjective responses were made after the initial *old/new* judgments. As a result, these judgments were only made for items that were not monitored (i.e., edited out) via the modality-related processes discussed earlier, although modality may have affected the subjective attributions. A 2 (study modality) \times 2 (list type) ANOVA revealed a main effect of list type [$F(1,30) = 7.00$, $MS_e = .122$] indicating that remember judgments to critical lures from associative lists ($M = .42$) were more frequent than remember judgments to critical lures from categorized lists ($M = .21$). There was no main effect of modality [$F(1,30) = 1.22$, $MS_e = .097$, $p > .27$], but there was a significant interaction between modality and list type [$F(1,30) = 5.92$, $MS_e = .096$]. This interaction suggests that modality affected remember judgments for associative lists and know judgments for categorized lists, potentially because associative list false recognition yields more remember responses in general, whereas categorized list false recognition yields more know responses. However, follow-up *t* tests revealed that neither of these effects was significant, so this finding is only tenuous. For remember judgments to list items, a 2 (study modality) \times 2 (list type) ANOVA revealed a main effect of modality [$F(1,39) = 7.82$, $MS_e = .026$]. Participants were more likely to give remember judgments to items studied from visually presented lists than they were to auditorily studied

list items, consistent with the idea that visual presentation was more distinctive. Remember judgments did not differ as a function of list type [$F(1,39) = 2.49, MS_e = .288, p > .10$], nor was there an interaction between modality and list type [$F(1,39) < 1$].

The main finding in Experiment 1 was an equivalent modality effect for associative and categorized lists. The modality effect observed for associative lists (11%) was similar to the 12% effect found by Gallo et al. (2001, Experiment 2). For categorized lists, we found a modality effect of equal magnitude (12%). This finding is inconsistent with a study-based monitoring process as the primary mechanism underlying the modality effect, at least as it applies to categorized lists. For these lists, monitoring the activation of critical items during study should be relatively infrequent, given that the activation of these items at study is known to be less frequent than in associative lists (S. M. Smith et al., 2002).

The equivalent modality effect obtained for associative and categorized lists suggests that when critical items were encountered during test, participants were better able to use monitoring processes to reduce false recognition after visual study than after auditory study. However, our findings are also consistent with the hypothesis that greater relational processing of auditorily presented lists relative to visual presentation resulted in higher false recognition of related lures. We tested these two accounts (test-based monitoring vs. relational processing) in Experiment 2.

EXPERIMENT 2

In Experiment 2, we substituted *meaning* instructions at test (see, e.g., Brainerd & Reyna, 1998; Schacter et al., 2001) for the standard *old/new* instructions used in Experiment 1. Meaning instructions require participants to remember the themes of lists they have studied rather than specific items, and should provide a measure of the relative amount of thematic gist that is cued by critical lures from the two types of lists. Arndt and Reder (2003) suggested that a relational processing explanation, rather than one based on decision or monitoring processes, could explain the modality effect. In their view, visual presentation of study items promotes greater item-specific processing, and less relational processing, than does auditory presentation. As a result, the modality effect may stem from differences in the encoding of relational information between the two study modalities. According to this explanation, if visual presentation leads to reduced levels of relational processing, then the encoding of gist-based information will be reduced, which should be reflected on a meaning-based test. That is, if the modality effect is due to different levels of thematic gist for the critical item, then a modality effect should be especially pronounced on a gist test.

In addition, the use of meaning instructions should render false recognition monitoring processes during test unnecessary. This is because participants were instructed to say “yes” to related words, regardless of whether the words were studied. If the modality effect is due to moni-

toring or decision processes during test, then it should be eliminated on a meaning-based test.

Method

Participants. A total of 40 volunteers participated in the experiment and were recruited from the same sources as in Experiment 1.

Design and Procedure. The study phase was identical to that in Experiment 1. At test, participants were given meaning instructions instead of *old/new* instructions. They were instructed to indicate whether each item was similar to one of the themes or concepts from the lists they had studied or whether it was dissimilar to any of the previously studied themes. Importantly, we stressed that it was irrelevant to their decision whether or not they remembered the test word from the study phase:

You will now be given a similarity test for the 24 lists that you studied. You may have noticed when studying these lists that each was organized around a central concept or theme. On the following test, a series of words will be presented. Some were studied and some were not, but we are not interested in your memory for these words. Instead, we want to know whether you think the test word is similar to a central concept or theme that was studied. Some studied and nonstudied words will seem similar to one of the themes, and others will not. On this test, regardless of whether you think a word was presented, we want you to press the key labeled “S” (for Similar) for any word that is similar to the concept or theme of any of the lists. We want you to press the key labeled “D” (for Dissimilar) for any word that is dissimilar to the concept or theme of any of the lists.

For example, suppose you were presented a list like “minute,” “second,” “year,” “day,” and “century.” If, on the test, you are presented with “minute” or “hour,” you would respond “S.” However, if presented with “gallon,” you would respond “D.” In another example, suppose you were presented a list like “boy,” “dolls,” “female,” “young,” and “pretty.” If, on the test, you are presented with “boy” or “girl,” you would respond “S.” However, if presented with “computer,” you would respond “D.”

As before, the participants were instructed to go through the test items at their own pace.

Results and Discussion

Table 2 shows the proportion of *similar* responses to list items, true target controls from nonstudied lists, critical items, and false target controls from nonstudied lists as a function of presentation modality and list type. As in Experiment 1, we analyzed raw and corrected scores, and we report only corrected results unless there were differences.

Table 2
Experiment 2: Proportions of List Items and Critical Items Judged as Similar to Studied Themes, as a Function of Study Modality of the Lists and List Type

	List Type	
	Categorized	Associative
Critical Items		
Auditory lists	.84	.90
Visual lists	.88	.93
Nonstudied lists	.16	.17
Corrected		
Auditory lists	.68	.73
Visual lists	.72	.76
List Items		
Auditory lists	.87	.81
Visual lists	.90	.81
Nonstudied lists	.12	.16
Corrected		
Auditory lists	.74	.65
Visual lists	.77	.65

A 2 (study modality) \times 2 (list type) ANOVA on corrected scores for critical items revealed no main effect of modality [$F(1,39) = 1.64$, $MS_e = .034$, $p > .20$] and no list type \times modality interaction [$F(1,39) < 1$]. There was a marginal effect of list type [$F(1,39) = 3.15$, $MS_e = .027$, $p = .084$] [this effect was significant on raw scores, $F(1,39) = 9.84$, $MS_e = .012$]. *Similar* responses to critical lures from associative lists were more frequent than those to lures from categorized lists, although this effect was small (5%). Although there was no modality effect on corrected *similar* responses, inspection of Table 2 shows that these responses on an uncorrected basis may have been subject to ceiling effects. Therefore, we identified a subset of 10 lists that were below ceiling ($M = .77$) and examined modality effects for these lists. Results showed an identical level of *similar* responses to these lists, whether the lists had been presented auditorily or visually (both $M_s = .77$).

As was the case with critical items, a 2 (study modality) \times 2 (list type) ANOVA on list items revealed no effect of modality [$F(1,39) < 1$] and no interaction [$F(1,39) = 1.94$, $MS_e = .007$, $p > .17$]. There was a main effect of list type [$F(1,39) = 26.30$, $MS_e = .018$] indicating that studied items from categorized lists were judged as similar to list themes significantly more often than were such items from associative lists. This main effect suggests that categorized list words cue their gist more than do associative list words, although differences in true memory (as observed in Experiment 1) may also have influenced these results.

To summarize the results of Experiment 2, we found that the use of meaning instructions at test eliminated the modality effect for both categorized and associative lists. This finding further supports the hypothesis that test-based monitoring underlies the modality effect, rather than different levels of relational processing arising from visual versus auditory study. That is, if the modality effect were due to monitoring, then it should be eliminated on the meaning-based or gist test (which does not require such monitoring); if it were due to different levels of gist, it should be augmented on the gist test.

GENERAL DISCUSSION

To summarize, we found an effect of presentation modality on false recognition from categorized lists that was equivalent to that observed from associative lists (Experiment 1). Because previous research has shown that activation of critical lures during study is lower for categorized lists than for associative lists (S. M. Smith et al., 2002), this finding suggests that the modality effect is driven primarily by a monitoring process that operates during test. That is, relatively low levels of conscious generation of related lures during study of these lists precludes the need for study-based monitoring, suggesting that the categorized list modality effect is due to test-based monitoring. By extension, we can infer that test-based monitoring also plays a role for associative lists, because there was no reason to expect it would differ for the two list types.

The use of meaning instructions in Experiment 2 allowed us to further examine the test-based monitoring hypothesis,

as well as test the monitoring versus relational processing account of the modality effect. These test instructions require participants to remember list themes rather than specific items, and provide a measure of thematic gist cued by both list items and critical items. By focusing participants on thematic information (i.e., gist) and instructing them to disregard whether or not the item was studied, we eliminated the need for monitoring during test. Consistent with the test-based monitoring hypothesis, the meaning instructions eliminated the modality effect for both lists.

By eliminating the modality effect in Experiment 2, we also provided evidence against the idea that the modality effect (for both lists) was due to different levels of gist or relational processing across modalities (Arndt & Reder, 2003). According to Arndt and Reder, auditory presentation results in greater relational processing (i.e., more gist) than does visual presentation (which instead encourages item-specific processing). Consequently, if the modality effect were due to different levels of gist-based processing across modalities, then it should have been preserved or been even greater on a gist test.

Our results extend those previously reported by Schacter et al. (2001) who found that a meaning test greatly attenuates reductions in false recognition on a standard old/new test that are observed after picture encoding as opposed to word encoding of associative lists. Schacter et al. argued that reduced false recognition after picture encoding on the standard test is attributable to a monitoring strategy known as the distinctiveness heuristic (Schacter et al., 1999; for a review, see Schacter & Wiseman, in press). On a meaning test, this monitoring strategy is not used, hence attenuating the picture/word encoding effects observed on the standard test, much as the modality effect was eliminated on our meaning test.

Although our focus in this study was on the modality effect in false recognition, several of our other findings also warrant discussion. In Experiment 1, we found that whereas categorized list items were recognized more frequently than associative list items, overall levels of false recognition, as well as remember judgments to critical lures, were substantially higher for associative lists than for categorized lists. These results are consistent with the finding that critical lures from associative lists are activated more frequently than lures from categorized lists (S. M. Smith et al., 2002). However, these list differences also are consistent with the idea that lures from associative lists may be better cues for thematic gist than are lures from categorized lists. We also found list differences on the meaning-based test in Experiment 2. Lures from associative lists were judged to be similar to list themes more often than were lures from categorized lists. This finding supports the idea that associative list lures are better cues for their gist, although it is also possible that enhanced associative activation of DRM (i.e., associative) lures made them more familiar, and this familiarity increased similarity judgments. Our study did not allow us to distinguish between these two explanations.

In conclusion, our findings are in general agreement with an activation/monitoring theory of false memory (see,

e.g., Roediger, Balota, & Watson, 2001; Roediger, Watson, et al., 2001). This theory proposes that activation of related lures, as well as processes involved in the monitoring of memory accuracy, determines levels of false memories. As discussed in the introduction, monitoring can occur both during study and during test. By finding similar modality effects for both categorized and associative lists, we showed that although associative activation may differ substantially between the different list types, similar monitoring processes may be engaged in the subsequent reduction of false memories. Our findings showing that test-based monitoring can reduce false memories have implications for minimizing memory distortion in real world settings. In those cases (e.g., eyewitness testimony), memory retrieval occurs long after the encoding phase, meaning that manipulations that can reduce false memories during the testing phase are vital. Such examples highlight the importance of examining further the role of monitoring processes that contribute to memory accuracy.

REFERENCES

- ARNDT, J., & REDER, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory & Language*, *48*, 1-15.
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, *80*(3, Pt. 2), 1-46.
- BRAINERD, C. J., & REYNA, V. F. (1998). When things that were never experienced are easier to "remember" than things that were. *Psychological Science*, *9*, 484-489.
- BRAINERD, C. J., WRIGHT, R., REYNA, V. F., & PAYNE, D. G. (2002). Dual-retrieval processes in free and associative recall. *Journal of Memory & Language*, *46*, 120-152.
- DEESE, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17-22.
- DODD, M. D., SHEARD, E. D., & MACLEOD, C. M. (2006). Re-exposure to studied items at test does not influence false recognition. *Memory*, *14*, 115-126.
- DODSON, C. S., & SCHACTER, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155-161.
- DODSON, C. S., & SCHACTER, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory & Language*, *46*, 782-803.
- EINSTEIN, G. O., & HUNT, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 588-598.
- GALLO, D. A. (in press). *Associative illusions of memory: False memory research in DRM and related tasks*. Philadelphia: Psychology Press.
- GALLO, D. A., McDERMOTT, K. B., PERCER, J. M., & ROEDIGER, H. L., III (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 339-353.
- GALLO, D. A., & ROEDIGER, H. L., III (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory & Language*, *47*, 469-497.
- HUNT, R. R., & EINSTEIN, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior*, *20*, 497-514.
- HUNT, R. R., & McDANIEL, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory & Language*, *32*, 421-445.
- ISRAEL, L., & SCHACTER, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*, 577-581.
- JOHNSON, M. K., HASHTRUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.
- JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67-85.
- KELLOGG, R. T. (2001). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 913-919.
- KORIAT, A., & GOLDSMITH, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490-517.
- MARSH, E. J., McDERMOTT, K. B., & ROEDIGER, H. L., III (2004). Does test-induced priming play a role in the creation of false memories? *Memory*, *12*, 44-55.
- 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., & WATSON, J. M. (2001). The rise and fall of false recall: The impact of presentation. *Journal of Memory & Language*, *45*, 160-176.
- McEVOY, C. L., NELSON, D. L., & KOMATSU, T. (1999). What is the connection between true and false memories? The differential roles of inter-item associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 1177-1194.
- RAJARAM, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*, 89-102.
- ROEDIGER, H. L., III, BALOTA, D. A., & WATSON, J. M. (2001). Spreading activation and the arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (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, & Cognition*, *21*, 803-814.
- ROEDIGER, H. L., III, & McDERMOTT, K. B., PISONI, D. B., & GALLO, D. A. (2004). Illusory recollection of voices. *Memory*, *12*, 586-602.
- ROEDIGER, H. L., III, WATSON, J. M., McDERMOTT, K. B., & GALLO, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385-407.
- SCHACTER, D. L., CENDAN, D. L., DODSON, C. S., & CLIFFORD, E. R. (2001). Retrieval conditions and false recognition: Testing the distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 827-833.
- SCHACTER, D. L., ISRAEL, L., & RACINE, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory & Language*, *40*, 1-24.
- SCHACTER, D. L., NORMAN, K. A., & KOUTSTAAL, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, *49*, 289-318.
- SCHACTER, D. L., & SLOTNICK, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron*, *44*, 149-160.
- SCHACTER, D. L., & WISEMAN, A. L. (in press). Reducing memory errors: The distinctiveness heuristic. In R. R. Hunt & J. Worthen (Eds.), *Distinctiveness and memory*. New York: Oxford University Press.
- SEAMON, J. G., LUO, C. R., SCHLEGEL, S. E., GREENE, S. E., & GOLDBERG, A. B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory & Language*, *42*, 120-146.
- SMITH, R. E., & HUNT, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, *5*, 710-715.
- SMITH, S. M., GERKENS, D. R., PIERCE, B. H., & CHOI, H. (2002). The roles of associative responses at study and semantically guided recollection at test in false memory: The Kirkpatrick and Deese hypotheses. *Journal of Memory & Language*, *47*, 436-447.
- UNDERWOOD, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122-129.

(Manuscript received May 26, 2004;
revision accepted for publication December 20, 2004.)