Reducing Gist-Based False Recognition in Older Adults: Encoding and Retrieval Manipulations

Wilma Koutstaal, Daniel L. Schacter, Lissa Galluccio, and Kathryn A. Stofer
Harvard University

Using a categorized pictures paradigm, Koutstaal and Schacter (1997) reported high levels of false recognition of lures that were categorically related to presented items. Although also shown by younger adults, false recognition was markedly higher for older adults. To probe the factors underlying this age difference, these experiments required participants to engage in more careful scrutiny of the items at retrieval or to notice specific differentiating perceptual features of the objects during encoding. False recognition was reduced with each of these manipulations, but neither manipulation, either separately or together, eliminated the age difference in false recognition. Older adults can considerably reduce false recognition if encouraged to use more stringent decision criteria. Persistent difficulty in opposing familiarity-based responding and comparatively more generic encoding may contribute to residual deficits.

Any true-to-life characterization of human memory must begin with two points of emphasis. On the one side, there is the general faithfulness of memory: Often reliable and accurate, memory is both a valuable and a necessary guide to behavior and judgment. On the other side, however, there is memory's susceptibility to error, including both negative errors or errors of omission (as when we fail to recall events or ideas or recollect them only vaguely) and positive errors or errors of commission. One may claim—sometimes with a high degree of confidence—that events or items that were never encountered were encountered previously or otherwise mistakenly attribute characteristics of an event that one did experience to another event, misaligning details of person, place, time, or other aspects of one's experiences (Estes, 1997; Roediger, 1996; Schacter, 1995; Schacter, Norman, & Koutstaal, 1998). Moreover, depending on circumstances, such errors of commission may prove to be more frequent among older than younger adults (Johnson, Hashtroudi, & Lindsay, 1993; Rankin & Kausler, 1979; Schacter, Koutstaal, & Norman, 1997; Spencer & Raz, 1995).

One form of positive memory error that has recently been subjected to considerable investigation among younger adults— and also, to a lesser degree, among older adults—is false recognition. Involving a mistaken claim or judgment that a novel word, object, or other stimulus was previously experienced, false recognition occurs when new items are in some way related to a studied item (Underwood, 1965). A number of recent studies have dem-

onstrated high rates of false recognition when using a paradigm in which numerous verbal associates, all of which converge on a single, nonpresented theme word are studied (Deese, 1959; Roediger & McDermott, 1995). Under these conditions, the nonpresented theme word later becomes an extremely difficult to resist *lure* item. Younger adults show surprisingly high rates of false recognition of the lures, with false recognition rates often approaching those of correct recognition (Roediger & McDermott, 1995; also cf. Mather, Henkel, & Johnson, 1997; Payne, Elie, Blackwell, & Neuschatz, 1996; Robinson & Roediger, 1997; Schacter, Verfaellie, & Pradere, 1996).

More recent extensions of this paradigm to older adults have shown that older adults show rates of false recognition that at least equal, or in some cases exceed, those of younger adults (Norman & Schacter, 1997; Tun, Wingfield, Rosen, & Blanchard, 1998). Older adults also may show elevated rates of false recognition for associates that are not highly associated to the theme word (Tun et al., 1998), respond comparatively more quickly when falsely recognizing items (Tun et al., 1998) and, in tests of free recall, show rates of intrusions of the nonpresented lure words that equal or surpass those shown by younger adults (Norman & Schacter, 1997; Tun et al., 1998). Note, however, that the overall magnitude of age differences in the susceptibility to false recognition in the converging associates paradigm is typically small and is not always observed (cf. Norman & Schacter, 1997, Experiment 1; Tun et al., 1998, Experiment 1).

In contrast to these relatively small effects, Koutstaal and Schacter (1997) have recently reported extremely large age differences in susceptibility to false recognition in a categorized pictures paradigm. In this paradigm, participants were required to make yes–no recognition decisions regarding whether they had or had not previously been shown detailed colored pictures of individual objects. For some of the recognition test items, participants had earlier been shown a large number of similar pictures from the same object category (e.g., chairs, cats, or teddy bears); for other items, they had been shown relatively few similar pictures. Consistent with other investigations in the verbal domain (Hintzman,

1988; Robinson & Roediger, 1997; Shiffrin, Huber, & Marinelli, 1995) and with earlier findings using pictorial stimuli (e.g., Bower & Glass, 1976; Strack & Bless, 1994), both older and younger adults showed higher rates of false recognition when many similar items had been encountered compared with when few or no categorically similar exemplars had been presented. However, older adults showed rates of false recognition that considerably exceeded those of the young and showed especially pronounced effects of the number of exemplars they had encountered. The average false recognition rate for the largest (18 item) categories for older adults across the three experiments was 64% compared with 29% for younger adults. For a category size of 9 items, the false recognition rate for older adults was 47% compared with 25% for younger adults.

Notably, this substantial age difference in false recognition was accompanied by two further consistent patterns. First, older and younger adults showed high—and essentially equivalent—levels of *correct* recognition for items for which numerous exemplars had been studied: Across three experiments, the average hit rate of older adults for the largest categories (18 items shown at study) was 81% compared with 82% for younger adults. Second, this absence of an age difference in hit rates for large category items was also accompanied by a notably different pattern of hits for categories in which only one item from the category had been studied. Here, older adults also showed an increased likelihood of errors of omission (i.e., misses) that younger adults did not. Whereas the hit rates of younger adults were largely unaffected by category size, the hit rates for one-of-a-kind object categories for older adults fell, on average, nearly 20% behind the hit rates shown by younger adults for these same categories and also considerably behind the hit rates shown by older adults themselves for items taken from categories in which many similar items had been studied.

The combination of all three of these results, including (a) equivalent hits by older and younger adults for studied items from large categories, (b) elevated false recognition among older adults for these same (large) categories, and (c) depressed correct recognition in older adults for one-of-a-kind items, suggested that older adults were relying on knowledge concerning the general kinds of items they had studied to a greater extent than younger adults. All three findings would be explained if older adults were especially influenced by the general perceptual or conceptual similarities of the items they had encountered: what has been called *gist* (Brainerd, Reyna, & Kneer, 1995; Reyna & Brainerd, 1995) or general similarity information (Curran, Schacter, Norman, & Galluccio, 1997; Hintzman, 1988; Hintzman & Curran, 1994, 1995). If, compared with their younger counterparts, older adults were especially relying on their knowledge of the general types or categories of objects they had studied to make recognition decisions, then their recognition of target items from categories in which many related items were studied should be particularly aided because these categories were quite salient at study. Furthermore, if younger adults did not rely on categorical information to the same extent, this might allow the hit rates of older adults to approach or match those of younger adults for these categories. However, reliance on such general similarities would not allow accurate differentiation between actually studied items (targets) and nonstudied but categorically related items (related lures or distractors), thus resulting in higher false recognition for older than

younger adults for large category lures. Finally, memory for the categorical nature of the items would also be expected to be less strong in instances in which only a single item from a category had been studied, and so reliance on gist-based representations here might more often fail to support correct recognition for older adults.

Taken together, these findings suggest that age-related increases in false recognition in this paradigm might reflect errors due to the general similarity of the distractor items to studied items (false alarms of older adults to items that were entirely novel were considerably less frequent than for categorically similar items). This observation of similarity-based errors in older adults is important because many instances of age-related increments in false recognition or recall may be accounted for as deriving from source confusions regarding the origins of information that had been presented or generated previously in the experiment; for example, confusing whether an object or event had actually been perceived or had only been imagined (G. Cohen & Faulkner, 1989; Henkel, Johnson, & DeLeonardis, 1998) or whether an event had been seen previously in a videotape or a photograph (Schacter, Koutstaal, Johnson, Gross, & Angell, 1997; for review, see Johnson et al., 1993; Schacter, Koutstaal, & Norman, 1997; Spencer & Raz, 1995). In contrast, because in the categorized pictures paradigm the new pictures were never presented or explicitly imagined, a simple source confusion in this strict sense cannot explain the results. The total pattern of findings suggests that older adults may also be more susceptible to similarity-based errors.

The aim of the experiments reported here was to further examine the basis of both correct and incorrect recognition responses of older compared with younger adults in the categorized pictures paradigm. A number of different interpretations of the apparently heightened level of "gist-based" responding among older adults are possible. One set of possibilities focuses on age-related differences in processes operative at encoding. It is possible that older adults attend to and encode fewer specific details of the pictures: possibly because they notice fewer perceptual details at study, or because they place comparatively greater emphasis on the object categories to which items belong. These possibilities are consistent with proposals of earlier investigators that older adults might encode verbal items in a less specific manner than younger adults (e.g., Craik & Rabinowitz, 1985; Hess, 1984; Rabinowitz & Ackerman, 1982; Rabinowitz, Craik, & Ackerman, 1982) and also with more recent proposals that memory errors in older adults may partially be due to the failure to encode differentiating information that would allow similar-but-not-identical items to be separately stored and accessed (Schacter, Norman, & Koutstaal, 1998). Another set of possibilities focuses on age-related differences in processes operative at retrieval. For example, it is possible that older adults simply used less stringent recognition decision criteria for items when many similar exemplars had been presented previously, possibly because they were less aware of the need for caution in designating similar-seeming items as previously encountered or because, in the absence of explicit instructions to the contrary, their default was to designate items as old if they generally fit with the types of items that they had studied or if they otherwise seemed to be familiar (cf. Norman & Schacter, 1997; Reder, Wible, & Martin, 1986; also see Schacter, Norman, & Koutstaal, 1998).

To explore these possibilities, we report three experiments that attempted to assess the role of encoding and retrieval processes in producing this pattern of apparently stronger-than-usual reliance on general similarity or gist information among older adults. Each of the experiments used the related pictures paradigm of Koutstaal and Schacter (1997) but with specific modifications aimed at reducing the differences in false recognition of older versus younger adults. Experiment 1 involved a manipulation of retrieval conditions, Experiment 2 involved a manipulation of encoding, and Experiment 3 involved altered conditions at both encoding and retrieval. The nature and rationale for each of these manipulations is further outlined in the following sections.

Experiment 1: Retrieval Manipulation

We began with an alteration of the instructions provided to participants at the time of retrieval, that is, during recognition testing. As noted previously, it is possible that older adults were simply less aware of the need for caution at the time of testing than were younger adults and thus more often falsely claimed to recognize new items that were similar to items they had studied. Although research in other paradigms has not always shown differences in response criteria among older compared with younger adults (Isingrini, Fontaine, Taconnat, & Duportal, 1995), more lenient responding by older adults has sometimes been found (Bartlett, Leslie, Tubbs, & Fulton, 1989). Moreover, signal detection analyses performed on the data from our previous studies (Koutstaal & Schacter, 1997) also provided some evidence that older adults were using more lenient criteria, especially for items for which many related exemplars had been presented.

The primary purpose of this experiment was to determine (a) whether providing instructions at the time of retrieval that discouraged designating items as "old" simply on the basis of general similarity to studied items would reduce gist-based false recognition, and (b) whether these instructions would prove particularly beneficial for older adults. To this end, at the time of test, rather than simply designating items as *old* (previously presented) versus *new* (not previously presented), as in a standard yes–no recognition paradigm, participants were asked to further distinguish between new items that were in some way similar to previously encountered items and those that were entirely new. Participants were asked to classify each test item as falling into one of three categories: items that were *old and identical* ("meaning it is exactly the same as one of the pictures you saw on your earlier visit"); *new but related* ("meaning it is a picture that was never previously presented but shares some characteristics of, or is in the same category as, an item you saw on your earlier visit"), or *new and unrelated* ("meaning it is a picture that has never been previously presented and is unlike any other picture presented on your earlier visit").

A manipulation that required participants to make decisions on an item-by-item basis regarding the relation of the test probe to studied items was selected rather than more general instructions regarding the categorical nature of the stimuli and a global admonishment regarding the need for caution, for two reasons: (a) It reinforces the necessity to make item-specific judgments on every trial, and (b) previous studies have shown that requiring participants to engage in this form of item-by-item decision monitoring is a particularly effective means of reducing memory errors. Among younger adults, such item-by-item monitoring was found

to reduce errors both in traditional source memory tasks (Dodson & Johnson, 1993) and in the misinformation paradigm (Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989). Importantly, it has also been found to reduce source misattribution errors (in the form of *false fame* judgments) in older adults (Multhaup, 1995; cf. Dywan & Jacoby, 1990). Extrapolating from these findings to our related pictures paradigm, the comparatively high rates of false recognition for categorically related items might similarly be reduced by the requirement for more careful item-by-item examination during testing. If the high levels of false recognition observed in the related pictures paradigm are, in part, due to the use of lenient response criteria at the time of testing, then the modified retrieval instructions should reduce false recognition. Furthermore, if overly lenient criteria are particularly problematic for older adults, then the modified retrieval instructions and test format should have an especially pronounced effect on false recognition of this age group, with older adults showing a greater reduction in false recognition (and probably also a greater shift toward more stringent response criteria) than younger adults.

Method

Participants. The experimental participants were 16 older adults (mean age = 69.4 years, range = 63–75 years) and 16 younger adults (mean age = 19.4 years, range = 17–24 years). The experimental participants were tested under exactly the same conditions as in Experiment 3 of Koutstaal and Schacter (1997); however, instead of performing a simple old–new judgment for the pictures, older and younger participants were asked to decide if the items were old and identical, new but related, or new and unrelated. The control participants were 16 older adults (mean age = 70.4 years, range = 64–75 years) and 16 younger adults (mean age = 19.7 years, range = 18–25 years); data from these participants have been reported separately by Koutstaal and Schacter (1997, Experiment 3). Older participants in both groups were recruited through newspaper advertisements and posted flyers and were individually interviewed so as to exclude those with a history of alcoholism or substance abuse, present or previous treatment for psychiatric illness, current treatment with psychoactive medication, drug toxicity, primary degenerative brain disorders, and brain damage sustained earlier from a known cause. Younger participants were recruited through posted sign-up sheets at Harvard University. Older adults in the experimental group had more years of formal education (M = 17.3, range = 13–22) than did younger adults in the experimental group (M = 13.4, range = 11–17), $F(1, 30) = 22.56, MSE = 5.41, p < .0001$; for controls, years of education for older adults (M = 14.4, range = 12–20) and younger adults (M = 13.9, range = 12–19) were not significantly different ($F < 1$). All participants had normal or corrected-to-normal vision and were paid for their involvement in the study.

Experimental design. The experimental design included two between-subjects variables of age (old or young), and retrieval condition (experimental or control). There was also a within-subjects variable of category size. For studied items, category size had three levels, with 1, 9, or 18 category exemplars presented. For nonstudied items, category size had four levels, with 0, 1, 9, or 18 related items presented at study. (False alarms to the zero condition provide a baseline measure of false recognition. False alarms to these items are referred to as *novel* false alarms because they entailed false alarms to categories that were novel at the time of test.)

Stimuli. The stimuli were detailed colored pictures of single objects (or, in a few cases, coherent groupings of objects), without background, that were taken from various illustrated books for children and adults. All pictures were mounted on white index cards and then scanned and converted to digital format using VistaScan and a UMAX Vista-S6E scanner. At both study and test, the pictures were displayed on a color computer

monitor, using a 256-color look-up table, and the PsyScope experimental presentation program.

The pictures portrayed objects from 20 different object categories, with each category comprising a total of 21 different exemplars. For example, there were 21 pictures of cars, 21 pictures of cats, 21 pictures of children, and so on. In addition, there were 30 pictures of unrelated objects (e.g., a painted wine jar and a unicycle), and additional categorized objects that were used as filler items (see later).

The 20 categories were first assigned to four equal sets of five categories each (P, Q, R, and S); for example, Set P included cars, cats, children, clocks, and flowers, and Set Q included birds, shelves, teapots, teddy bears, and whales. These four sets were then rotated through the four experimental conditions such that each set equally often served as nonstudied (or novel) items (presented only at test) or served as a study category comprising 1, 9, or 18 related items. When a given object category served as a large (18-exemplar) category, then all but 3 of the items were presented at study (the remaining 3 items were reserved to be presented during the recognition test as new but related items). Likewise, when a given category served as a medium (9-exemplar) or single (1-exemplar) category, only a subset of the total pool of items from that category was presented at study. In the latter cases, the particular items that were excluded was determined randomly, with the same items always excluded whenever an item comprised a 9-exemplar or 1-exemplar category.

As in the previous experiments, to avoid confounding the number of items per category that were presented at study with the number of items per category that were presented at test, each studied category (with one exception, noted later) was tested an equal number of times: 3 times with a studied item and 3 times with a lure item. This was accomplished by selecting a subset of items from each category, which always served as the critical study and test items. For each category, 6 items were initially randomly selected to serve as the critical target and lure items. These items were then assigned to two subsets (Subsets A and B) and were rotated through the study and test lists such that each subset equally often served as targets and lures for the studied categories or as novel items for the nonstudied categories. The novel categories were also tested three times.

The one exception concerned single-item categories. These categories were tested twice: once with the single presented study item (the target) and once with a related lure. For these purposes, one item from each of the two subsets (designated as A1 and B1) was randomly selected and rotated across the study–test conditions in the same manner as the 3-item sets.

Each study list comprised a total of 215 items, including 140 critical items (5 single-, 45 medium-, and 90 large-category exemplars), 54 filler items (two 9-item and two 18-item categories), 6 buffer items, and 15 unrelated items. Each test list comprised a total of 115 items, including 85 critical items, 15 studied unrelated items, and 15 unrelated new items. The filler and unrelated items were included to increase the variety and length of the study lists so as to maintain acceptable levels of recognition and also to match conditions of the earlier experiments; these items were not scored.

Procedure. The experimental procedure involved two main phases: a study phase and a test phase, separated by a retention interval of 3 days. Both the study and the test were administered individually.

In the study phase, participants were presented the pictures, one at a time, and were asked to rate their liking for each item. The pictures from different object categories were randomly intermixed (i.e., not blocked by category), and the encoding task was incidental; no mention was made of a subsequent memory test.[1] Each picture was presented for 2 s in the center of the computer monitor and was followed by a prompt, requesting participants to enter their liking rating (1 = *don't like*, 5 = *very much like*). Participants had as much time as they needed to enter their liking rating and then advanced to the next item by pressing the tab key. Following presentation of the last picture, participants were given an unrelated filler task for 5 min, and were then reminded of their following appointment.

In the test phase, participants were presented a subset of the pictures from the liking rating task together with new (nonstudied) pictures. For the

control condition, the test consisted of a simple old–new recognition test, in which participants first indicated whether each item was old (previously presented during the liking encoding task) or new (not previously presented during the experiment) and then rated their confidence in their recognition judgment on a 5-point scale (1 = *just guessing*, 5 = *very sure*). For participants in the experimental condition, the test consisted of exactly the same items but with additional instructions and the requirement of a three-level recognition decision rather than an old–new decision. Participants were instructed that they should indicate whether each picture was old and identical ("meaning it is exactly the same as one of the pictures you saw on your earlier visit"); new but related ("meaning it is a picture that was never previously presented but shares some characteristics of, or is in the same category as, an item you saw on your earlier visit"), or new and unrelated ("meaning it is a picture that has never been previously presented and is unlike any other picture presented on your earlier visit"). For items that they designated as new but related, participants orally stated the category to which the new item was believed to be related. Finally, participants provided confidence rating for all items. (In the interests of space, results relating to the confidence ratings are not presented in the text. However, Table 4 provides a summary of the high confidence responses for all three experiments.) The recognition test in both groups was self-paced. Following completion of the test, participants were debriefed.

Results

Figure 1 presents the proportion of correct and false recognition responses separately as a function of age (old vs. young) and category size (0, 1, 9, or 18 related exemplars presented at study). Correct responses are shown in the upper panel and false recognition responses are shown in the lower panel.

Table 1 shows the outcomes for the signal detection measures. We used the nonparametric signal detection measures of A' (a measure of sensitivity) and B_D'' (a measure of response bias) because values of A' have been shown to result in less error than values of d' under conditions in which performance shows some bias (Donaldson, 1993). Values of A' can vary between 0.00 and 1.00, with higher values indicating greater sensitivity and chance performance being .50; the corresponding bias measure B_D'' varies between −1.00 (extremely liberal) and +1.00 (extremely conservative), with 0 indicating unbiased responding (Grier, 1971; Hodos, 1970). Because these measures are undefined with hit rates of 0 or 1, the data were first transformed by computing $p(x)$ as $(x + .5)/n + 1$ rather than x/n (as recommended by Snodgrass & Corwin, 1988). In addition, in instances in which individuals showed below-chance sensitivity (hits < false alarms, or A' <

---

[1] Although participants were not told that their memory would be tested, it is possible that some individuals nonetheless anticipated a later test and so explicitly attempted to remember the pictures. On a post-experimental questionnaire that asked if participants were "specifically trying to remember the pictures," more older adults than younger adults responded "yes" (rates of 22% and 9% for old and young, respectively); similar outcomes were observed in Experiment 2 (30% and 19%, respectively; 5 observations missing), and Experiment 3 (50% and 25%, respectively). Comparing the results for participants who responded positively versus negatively to this question revealed no consistent patterns: In Experiment 1, older control participants who did not explicitly try to remember the pictures showed numerically higher false recognition than did older controls who reported trying to do so. However, no systematic pattern was found for the older experimental participants or for either group in Experiment 2, and the opposite pattern, with older *no* respondents showing lower false recognition than older *yes* respondents, was found in Experiment 3.
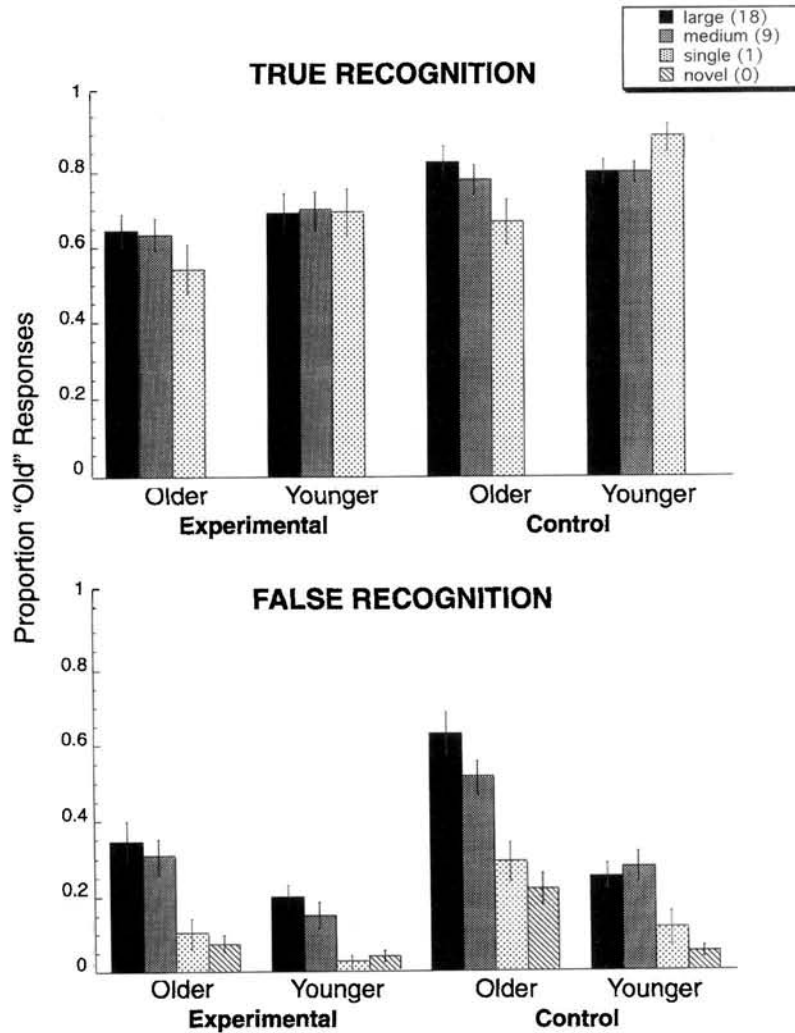
Figure 1. Mean proportion of old responses in Experiment 1 for studied items (true recognition) and nonstudied items (false recognition) as a function of category size, age, and retrieval condition. Category size, or the number of categorically related items presented during study, was 0, 1, 9, or 18 items; the category size of 0 provides the baseline false alarm rate. Results are shown separately for older and younger adults in the experimental group (given the tri-part retrieval test format) and the control group (given standard old–new instructions). Error bars show the standard error of the mean.

.50), the modified formulas for calculating sensitivity and bias for below-chance performance (Aaronson & Watts, 1987) were used. Measures of sensitivity and response bias were obtained, both comparing hits to novel false alarms (termed A'-novel and B_D''-novel, respectively, shown in the upper portion of Table 1) and comparing hits to related false alarms (A'-related and B_D''-related respectively, shown in the lower portion of Table 1).

False recognition and response bias. From Figure 1, it is clear that, as expected, false recognition was considerably reduced in the experimental compared with the control group. Combining across age and considering false recognition responses for all four types of new items (novel category lures, and single, medium, and large category lures), the rate of false recognition in the experimental condition (16%) was just over half that observed in the control condition (29%). A 2 × 2 × 4 (Age × Retrieval Condition ×

Category Size) analysis of variance (ANOVA) performed on the false recognition scores revealed a main effect of condition, $F(1, 60) = 26.57$, $MSE = .05$, $p < .0001$. Analyses conducted on the false recognition scores after correcting for differences in baseline levels of false alarms (i.e., false recognition responses minus novel-category false alarms) likewise yielded a main effect of condition, $F(1, 60) = 7.53$, $MSE = .04$, $p = .008$.

This across-age-group reduction in false recognition was also accompanied by alterations in response criteria. Separate 2 × 2 × 3 (Age × Retrieval Condition × Category Size) ANOVAs performed on the response bias measures of B_D''-novel and B_D''-related revealed main effects of retrieval condition on both indexes. Participants in the experimental group used more conservative criteria than those in the control group, both when hits were compared with novel false alarms (Ms = .63 and .22, respec-

Table 1

Measures of Sensitivity and Response Bias, Experiment 1: Retrieval Manipulation

	Experimental				Control			
	Old		Young		Old		Young	
Condition	A'	B''_D	A'	B''_D	A'	B''_D	A'	B''_D
Item-specific memory (hits compared to novel false alarms)								
Single								
M	.83	.67	.88	.66	.79	.24	.93	.32
SD	.07	.36	.07	.36	.12	.59	.04	.40
Medium								
M	.86	.60	.90	.58	.85	−.002	.92	.46
SD	.05	.42	.05	.55	.06	.58	.04	.40
Large								
M	.86	.62	.89	.63	.87	−.14	.92	.44
SD	.06	.35	.06	.37	.08	.55	.04	.44
Item-specific memory (hits compared to related false alarms)								
Single								
M	.78	.54	.86	.49	.73	.04	.89	−.03
SD	.11	.44	.08	.37	.14	.54	.08	.50
Medium								
M	.75	.11	.85	.26	.71	−.48	.83	−.16
SD	.07	.52	.08	.60	.10	.45	.08	.47
Large								
M	.73	.05	.81	.17	.67	−.59	.85	−.12
SD	.09	.59	.10	.53	.17	.48	.05	.47

Note. Category size (number of categorically related items presented): for single condition, 1 item per category; for medium condition, 9 items per category; for large condition, 18 items per category.

tively), $F(1, 60) = 15.45, MSE = .52, p = .0002$, and when hits were compared with within-category false alarms ($Ms = .27$ and $−.22$, respectively), $F(1, 60) = 24.13, MSE = .48, p < .0001$.

These overall effects of retrieval condition on false recognition and response criteria were also accompanied by some (albeit less strong and less consistent) evidence that older adults were particularly influenced by the retrieval manipulation. In the overall (uncorrected) analysis of false recognition responses, older adults showed a greater reduction in false recognition under the experimental relative to the control condition ($Ms = .21$ and $.41$, respectively) than did younger adults ($Ms = .10$ and $.17$, respectively), $F(1, 60) = 6.59, MSE = .05, p = .01$, for the Age × Retrieval Condition interaction. However, this differential pattern was no longer found after correcting for differences in the baseline rates of false alarms ($F < 1$ for the Age × Retrieval Condition interaction).

For the measures of response bias, older adults in the experimental group tended to show a greater shift toward increased stringency of responding on the B_D''-novel measure ($Ms = .03$ and .63 for control and experimental, respectively) than did younger adults ($Ms = .41$ and .62, respectively), $F(1, 60) = 3.42, MSE = .52, p = .07$, for the Age × Retrieval Condition interaction.[2] In addition, on this measure, older experimental participants (similar to the younger participants in both groups but unlike older controls) showed uniformly conservative criteria, regardless of the number of exemplars they had studied: $F(2, 120) = 4.66, MSE = .05, p = .01$, for the Age × Retrieval Condition × Category Size

interaction (means for single, medium, and large categories of .67, .60, and .62 for older experimentals compared with .24, −.002, and −.14 for older controls; see also Table 1). However, on the B_D''-related measure of response bias, there was no Age × Retrieval Condition interaction and also no three-way interaction ($Fs < 1$). Overall, both older and younger experimental participants showed more conservative responding than did participants in the control condition, but the criteria of older adults did not shift more than that of younger adults. Moreover, on this measure, older experimentals continued to be strongly influenced by category size in a manner parallel to that shown by the older controls (see Table 1).

Taken together, these findings indicate that the retrieval monitoring manipulation generally benefitted older adults as much—or in some cases more—than younger adults. Nonetheless, from Figure 1 it is also clear that the false recognition rate of older adults was not reduced to the same level as that of younger adults. Analyses confined to the experimental group alone indicated that older adults still showed significantly greater false recognition than did younger adults, both in absolute terms, combining across all four lure types (overall $Ms = .21$ and .10, respectively), $F(1, 30) = 8.83, MSE = .04, p = .006$, and following correction for novel false alarms, $F(1, 30) = 4.15, MSE = .04, p = .05$. Moreover, this age-related difference in false recognition was observed irrespective of the number of related exemplars that had been studied.[3] Separate pairwise comparisons showed that false recognition of the older experimental group exceeded that of younger experimentals for each of the three category sizes (1, 9, or 18 items; smallest $F = 6.06$).

Correct recognition (hits) and sensitivity. An initial 2 × 2 × 3 (Age × Retrieval Condition × Category Size) ANOVA performed on the uncorrected recognition scores showed a main effect of condition, with overall recognition in the experimental group ($M = .66$) substantially lower than recognition in the control group ($M = .80$), $F(1, 60) = 14.49, MSE = .07, p = .0003$. A similar trend toward reduced recognition in the experimental relative to the control condition was found after correcting for novel-category false alarms, $F(1, 60) = 3.12, MSE = .07, p = .08$. In neither

---

[2] For a nondirectional test, with alpha at .05, a power analysis showed that the likelihood of detecting this medium-small interaction effect ($f = .24$) was approximately 65% (J. Cohen, 1988; Rosenthal & Rosnow, 1991). Estimates of power for other Age × Retrieval Condition interactions that likewise showed nearly significant effects in this experiment ranged between 56 and 69%, falling somewhat below the recommended power level of 80% (J. Cohen, 1988). The modest power of our experimental design ($n = 16$ per Age × Retrieval Condition cell) should be factored into any interpretation of the results.

[3] Both older and younger adults, regardless of condition, showed substantially greater false recognition for the many-exemplar than for the single or novel categories (see Figure 1). Main effects of category size on false recognition and novel-corrected false recognition were consistently observed in each of the experiments reported here. However, in the interests of a more economical presentation, only interactions of category size with age or experimental condition are reported, with the latter reported only if there was also an Age × Category Size interaction in the experimental condition alone. Likewise, overall main effects of age are not reported, emphasis being placed on any interactions of age with other variables.

instance, however, was there an Age × Condition interaction ($F < 1$ for the uncorrected analysis; $F < 2.4$ for the novel-corrected measure). Furthermore, similar overall impairments were not seen on the measures of sensitivity ($Fs < 1$) for the main effects of condition for A'-novel and A'-related. These results suggest that although both age groups tended to respond more conservatively in the experimental than the control condition (resulting in depressed recognition), this shift in criteria was not accompanied by a cost in sensitivity. Indeed, these initial analyses also showed that older adults in the experimental condition demonstrated somewhat greater sensitivity than those in the control condition, both on the A'-novel measure ($Ms = .85$ and $.83$, respectively) and the A'-related measure ($Ms = .75$ and $.70$, respectively) whereas a tendency toward the opposite pattern was seen for younger adults ($Ms = .89$ vs $.92$, and $.84$ vs $.86$), $F(1, 60) = 3.12, MSE = .008$, $p = .08$, and $F(1, 60) = 3.63, MSE = .01, p = .06$, for the Age × Condition interactions.

Analyses confined to the experimental group alone showed no overall age-related impairment in recognition for the uncorrected recognition scores ($F < 2$) but significant impairment following correction for novel false alarms, $F(1, 30) = 4.59, MSE = .08$, $p = .04$. Both the A'-novel and A'-related measures of sensitivity revealed an overall impairment in sensitivity for older adults in the experimental group, $F(1, 30) = 6.03, MSE = .007, p = .02$, and $F(1, 30) = 17.27, MSE = .011, p = .0002$, respectively. Follow-up analyses indicated that, for the A'-related measure, these age-related impairments in sensitivity were significant for all three category sizes (smallest $F = 6.70$). For the A'-novel measure, recognition of the one-of-a-kind items was significantly depressed, $F(1, 30) = 5.62, MSE = .005, p = .02$, and there was a trend toward impairment for the 9-item categories, $F(1, 30) = 3.82, MSE = .003, p = .06$, but not for the 18-item categories ($F < 2.3$).

Discussion

The primary objective of this experiment was twofold: to determine if (a) the inclusion of instructions at the time of retrieval that discouraged designating items as old simply on the basis of general similarity to studied items would reduce gist-based false recognition; and (b) if so, whether these instructions would prove particularly beneficial for older adults. The results suggest that the answer to both of these questions is clearly "yes." Our findings indicate that (a) similarity-based false recognition responding by older adults can be substantially reduced through establishing conditions that require more attentive and careful responding at the time of retrieval; and (b) such conditions may considerably reduce age-related differences in response criteria, allowing older adults to modify their decision making toward more conservative responding. These findings are consistent with previous studies with younger adults that found reduced likelihood of source memory errors (Dodson & Johnson, 1993) and reduced susceptibility to misinformation (Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989) when participants made more specific source decisions for each item. These outcomes are also consistent with Multhaup's (1995) findings that older adults were less likely to wrongly attribute experimentally induced familiarity of names to their preexperimental "fame" when they were required to make more differentiated judgments, categorizing each name as famous, a nonfamous name presented earlier, or a new nonfamous name.

Because, relative to our earlier experiments, we held the encoding conditions constant, the reduction in false recognition must have derived from an alteration in the way older adults reached their recognition decisions; that is, the locus of the beneficial effects must have been at retrieval. Moreover, this reduction in false recognition was achieved at relatively little cost to sensitivity. Considering the simple pattern of hits, the absolute recognition rate in the experimental group for both older and younger adults was less than that for the control group, with the overall hit rate in the older experimental group down by 15% (relative to the controls) and the younger experimental group down by 14%. However, the measures of sensitivity did not reflect decreases in the experimental relative to the control group: There was no overall main effect of condition for either the measure of sensitivity comparing hits with novel false alarms (A'-novel) or comparing hits with related false alarms (A'-related). Indeed, older adults (though not younger adults) showed slightly greater sensitivity in the experimental than in the control condition.

It appears that the explicit and repeatedly reinforced requirement that participants differentiate old and identical items from new but related items encouraged participants to consider and evaluate the recognition test items more closely. For example, participants may have tried to determine if a given item possessed any specific perceptual features that they had also noticed during the study phase and that would serve to vouch for its status as an old and identical item. In the absence of such distinctive perceptual information—or other forms of information, such as the recollection of a specific affective, cognitive, or evaluative response—participants may have opted to "play it safe" and call the item new but related. These possibilities are similar to a proposal of Johnson and colleagues (Dodson & Johnson, 1993; Johnson et al., 1993; also cf. Mather, Henkel, & Johnson, 1997; Norman & Schacter, 1997; Schacter, Norman, & Koutstaal, 1998) that, compared to simple yes–no recognition, source monitoring tests may require participants to use different criteria—perhaps stricter, but also possibly qualitatively different—in providing their judgments. Rather than simply responding on the basis of the general familiarity of the items (or type of item), more fine-grained distinctions among targets and lures—such as required in source monitoring judgments or the old and identical, new but related, and new and unrelated judgments of our experiment—may encourage participants to require, and actively search for, specific qualitative information that may provide a "warrant" for the old and identical designation.

Thus, these across-group (experimental vs. control) comparisons clearly demonstrate the value of encouraging more stringent retrieval monitoring, with the recognition performance of both younger and older adults clearly enhanced. Nonetheless, this experiment also yielded several indications that older adults in the experimental condition were still more susceptible to error, including both errors of omission and errors of commission, than their younger counterparts. Older adults in the experimental condition achieved fewer hits (correct recognition responses) than did younger adults, with this difference numerically most pronounced for the one-of-a-kind items. The analyses of sensitivity also showed that older adults in the experimental group had significantly impaired sensitivity compared with younger adults, both

when hits were compared with novel false alarms and when hits were compared with related false alarms. In addition, although false recognition in the older experimental group was considerably reduced relative to that observed in the older control group, the retrieval manipulation did not eliminate age-related differences in false recognition. False recognition among older adults in the experimental group still significantly exceeded that of younger adults in the experimental group.

In combination, these findings suggest that although one variable contributing to older adults' increased level of false recognition involves differences in response criteria, this cannot account for all of the difference. Even when the two age groups showed similar response criteria as a result of an instructional set and test format that encouraged careful and stringent responding, older adults showed elevated rates of false recognition for categorically related lures and decreased rates of correct recognition. This suggests that a further variable (or combination of variables) may be important. We next consider one such possibility: the extent to which, and the manner in which, older adults initially encoded the stimuli.

Experiment 2: Encoding Manipulation

The possibility that older adults might encode items in a less specific or more generic manner than younger adults has been considered by a number of researchers (Hess, 1990; Isingrini et al., 1995; Labouvie-Vief, 1980; Rabinowitz & Ackerman, 1982; Rabinowitz et al., 1982). Some support for this hypothesis has been obtained, particularly from paradigms that evaluate the degree of benefit that older versus younger adults achieve from the reinstatement of specific features of the encoding context (Hess & Higgins, 1983; Rabinowitz & Ackerman, 1982; but also see Naveh-Benjamin & Craik, 1995; Schramke & Bauer, 1997), with observations that older adults are less adversely affected by altered context than younger adults, interpreted as reflecting a relative lack of encoding specificity (Tulving & Thomson, 1973). However, in the related pictures paradigm, increased specificity of encoding of the pictures does not necessarily require more extensive encoding of external contextual or semantic features but, instead, would appear to require increased attentional focus on features *within a single perceptual stimulus* (though, note that to be useful these must also differentiate an object within a class from other objects within the class; cf. Schacter, Norman, & Koutstaal, 1998). Furthermore, the encoding conditions in Experiment 1 and the earlier experiments were already, at least to some extent, conducive to distinctive encoding. Both the nature of the stimulus materials (highly detailed color pictures) and the orienting task that was used (rating the degree to which they liked the pictures) should have encouraged item-specific processing (Einstein & Hunt, 1980; Weldon & Roediger, 1987; also see Dewhurst & Conway, 1994; Park, Puglisi, & Smith, 1986).

Nonetheless, it is possible that younger and older adults responded somewhat differently to this encoding task, or that age differences in encoding might be reduced with an orienting task that provided more extensive support to extract distinctiveness. For example, specific instructions to review factual aspects of an experience may reduce, or even eliminate, the frequently observed age-related deficits in the ability to identify the source of information (Hashtroudi, Johnson, Vnek, & Ferguson, 1994). More-

over, although the instructions for the encoding task included specific instructions to participants that they should make their liking ratings on the basis of the individual items presented rather than the general class or type of thing to which the object belonged, it is possible that older adults more often made their liking ratings on the basis of the general class of the items or were, for other reasons, more focused on the categorical nature of the items, and less on item-specific information, than were younger adults. Research with younger adults (Marks, 1991) has shown that participants who were encouraged to adopt a categorical orientation to pictures (line drawings of scenes) showed higher rates of false recognition than did participants who were encouraged to focus on physical features of the objects. These observations suggest that, to the extent that older adults were more likely to process the pictures in a categorical manner, false recognition may have been increased.

The aim of Experiment 2 was to examine performance of older and younger adults in the related pictures paradigm when greater support for the extraction of distinctive perceptual features of individual items was provided at encoding. Immediately before the presentation of each object, participants in the experimental condition were provided with a brief verbal description that pointed to two relatively distinctive or individuating features of the object. Participants were asked to look for and notice these features during the subsequent presentation of the object. For example, one of the pictures was preceded by a prompt to "Notice the compact, pudgy body and round feet of this teddy bear"; other pictures were preceded by similarly specific descriptors: "Notice the silvery-white color and attached seat cover of this motorcycle"; "Notice the long, thin neck and spiked crest of this bird," and so on.

The false recognition of older and younger adults in this experimental condition was contrasted with that of older and younger adults in a control condition that was similar to the experimental condition but that did not provide distinctive information (see description of stimulus materials in the following section). If one of the variables contributing to false recognition of related items in this paradigm involves insufficient attention to and encoding of perceptual features that differentiate various exemplars from one another, then the provision of these verbal descriptions should decrease overall false recognition. Furthermore, if older adults were particularly likely to rely on overly general categorical information about the items or otherwise limited the extent to which they perceptually processed the items, then—to the extent that these verbal descriptions encourage more attentive perceptual processing—false recognition errors of older adults should be differentially reduced.

These predictions focus on the likely effects of the manipulation on false recognition; the probable outcomes of the encoding task for veridical recognition, particularly among older adults, are less clear. On the one hand, directing participants' attention to specific perceptual details of the objects should enhance item-specific memory, thereby increasing correct recognition. On the other hand, if the encoding task alters participants' willingness to designate items as old simply on the basis of broad categorical similarities or gist, leading them to require something more of each item before classifying it as old, then this could act to reduce gist-based *correct* recognition (not only false recognition) and thus offset any gains achieved by the increased encoding of item-specific information. Furthermore, to the extent that older adults

were previously more reliant on gist than were younger adults—as suggested by their contrasting correct recognition performance for many-exemplar items compared with one-of-a-kind items—this could differentially reduce older adults' correct recognition.

Method

Participants. The participants were 32 older adults (*M* age = 67.7 years, range = 61–74 years) and 32 younger adults (*M* age = 20.1 years, range = 16–27 years) recruited, screened, and reimbursed in the same manner as in Experiment 1. All participants had normal or corrected-to-normal vision and were native speakers of English. Older adults had, on average, more years of formal education (*M* = 15.6, range = 12–24) than younger adults (*M* = 14.0, range = 11–19), $F(1, 60) = 5.71, MSE = 6.84$, $p = .02$. This difference was true for both the experimental and control groups (*F* < 1 for the main effect of condition and for the Age × Condition interaction).

Experimental design. The experimental design included two between-subjects variables of age (old, young) and encoding condition (experimental, with descriptions; control, with filler task). There was also a within-subjects variable of category size. For studied items, category size had three levels: 1, 9, or 18 items presented. For nonstudied items, category size had four levels: 0, 1, 9, or 18 related items presented at study.

Stimuli. The pictorial stimuli were identical to those used in Experiment 1 except that to keep the total study phase duration to a reasonable time period, the 54 nontested filler items that were previously included in the study phase were omitted.

In the experimental group, all pictures were preceded by a brief verbal description that pointed to two distinctive features or characteristics of the subsequent pictured object and that also provided the name of the category to which the object belonged (e.g., "Notice the smooth black hair and big gold eyes of this cat"). Participants in the control condition were also exposed to the name of the category to which the object belonged. However, rather than reading a description of features regarding the immediately following object, the category name in the control condition was preceded by a short string of mixed letters and other characters, such as "i—idgfrxwiwfigtefa—CAT" or "4—835682464876—FLOWER." Participants in the control condition were asked to read the first set-aside character (here, "i" and "4," respectively) and then scan the following string for occurrences of that same character, counting out loud each time a repetition of the character was encountered, until they reached the end of the string, when they were to also read (out loud) the category name of the following item. Thus, both the experimental and control groups performed tasks in the interval between pictures, and both groups were presented the category labels of the pictures before the pictures themselves. However, only the experimental group received information designed to encourage careful perceptual processing of the items.

The generation of the distinctive descriptions for the pictures occurred in two phases. First, several younger adults were shown all of the pictures in a given object category simultaneously, with index cards, and were asked to provide two characteristics of each object that, together, would uniquely identify the object from all of the others in the category. However, this proved to be a demanding task, and many of the descriptions given did not include such unique identifiers. Thus, in the second phase, two of the experimenters (LG and WK) jointly evaluated the characteristics that were provided by the younger participants and then supplemented or revised these characteristics until both of the experimenters agreed that, together, the two characteristics provided a unique descriptor of the object relative to the other objects in the category.

The descriptions were presented in sentence format, with each sentence assuming the following form: "Notice the (*characteristic #1*) and (*characteristic #2*) of this (*object category name*)." The text was presented in the center of the computer screen in 18-point Geneva font on three separate lines, with the uppermost line containing the sentence beginning, the middle line providing the two characteristics, and the final line providing the object category name.

Procedure. The experimental procedure was, in broad outline, similar to that for Experiment 1. However, several changes were made in relation to the encoding phase: (a) The presentation of each item was preceded by either a verbal description (including the object category name) or a brief filler task followed by the category name of the following object. Participants in the experimental condition were instructed to read the description out loud and remember it for the upcoming picture. If they noticed the relevant details in the picture they were to press the "v" key on the computer keyboard; if they did not notice the specified details they were not to press the key. Participants in the control condition performed the filler task (described earlier) and then verbally stated their degree of liking for the picture as it was displayed on the screen (using the terms *like, don't like,* or *neutral*). The descriptions or filler task were each presented for 9 s. (b) To provide participants with sufficient time to notice and respond to the relevant distinctive features (experimental group) or to provide their verbal liking ratings (control group), the stimulus exposure time of the pictures was increased to 4 s (rather than 2 s). (c) Given the modifications of the encoding task as well as the increased stimulus exposure time, so as to avoid ceiling effects in recognition the retention interval between study and test was increased to 1 week (rather than 3 days as in the previous experiments).

Results

The proportions of correct and false recognition responses are shown in Figure 2. Table 2 presents the signal detection measures.

False recognition and response bias. An initial 2 × 2 × 4 (Age × Encoding Condition × Category Size) ANOVA performed on the false recognition responses revealed, as expected, a main effect of condition, with the overall level of false recognition in the experimental condition (*M* = .24) significantly lower than in the control condition (*M* = .30), $F(1, 60), = 5.40, MSE = .05$, $p = .02$. A similar pattern was observed following correction for baseline differences in novel-category false alarms, $F(1, 60) = 5.42, MSE = .0002, p = .02$. Considering the overall level of false recognition, this reduction was greater for older adults (*Ms* = .28 and .41) than for younger adults (*Ms* = .20 for both conditions), $F(1, 60) = 4.73, MSE = .05, p = .03$, for the Age × Condition interaction. However, this differential pattern was no longer apparent after correcting for novel-category false alarms (*F* < 1).

Analyses of the response criteria measures were also consistent with reduced gist-based responding. Participants in the experimental condition were significantly more conservative than those in the control condition, both when comparing hits to novel false alarms (*Ms* = .48 and .25, respectively), $F(1, 60) = 3.86, MSE = .68, p = .05$, and when comparing hits to related false alarms (*Ms* = −.003 and −.30, respectively), $F(1, 60) = 9.37, MSE = .46, p = .003$. In addition, on the B_D''-novel measure, there was a trend toward an interaction of age with condition, reflecting a somewhat greater shift toward conservatism among older adults (*Ms* = .50 and .07) than younger adults (*Ms* = .46 and .43), $F(1, 60) = 2.77, MSE = .68, p = .10$. For the B_D''-related measure, although older adults likewise showed a numerically greater shift (*Ms* = −.09 and −.50) than did
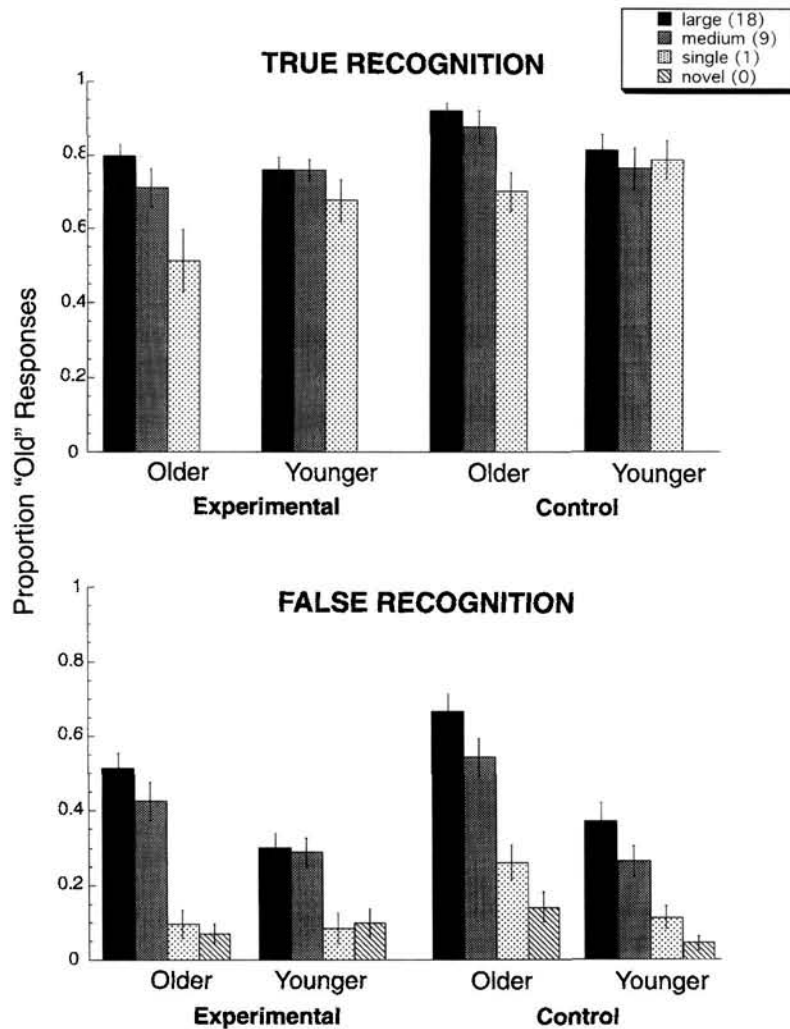
Figure 2. Mean proportion of old responses in Experiment 2 for studied items (true recognition) and nonstudied items (false recognition) as a function of category size, age, and encoding condition. Category size, or the number of categorically related items presented during study, was 0, 1, 9, or 18 items; the category size of 0 provides the baseline false alarm rate. Results are shown separately for older and younger adults in the experimental group (given the perceptually focused verbal descriptions at encoding) and the control group (given the filler and liking rating tasks at encoding). Error bars show the standard error of the mean.

younger adults ($Ms = .08$ and $-.11$), the interaction was not significant ($F < 1.3$).[4]

These outcomes indicate that older adults derived at least as much, and in some cases more, benefit from the encoding manipulation than younger adults. Nonetheless, analyses confined to the experimental condition alone revealed that several age-related differences remained. Relative to that observed for younger adults, older experimental participants still showed significantly elevated false recognition, both across all four category size conditions, ($Ms = .28$ and $.20$ for old and young, respectively), $F(1, 30) = 4.96$, $MSE = .05$, $p = .03$, and following correction for differences in baseline false alarms, $F(1, 30) = 13.89$, $MSE = .04$, $p = .0008$. In addition, for both analyses, there were significant interactions of age with category size, with the false recognition responses of older experimentals more influenced by category size

than their younger counterparts ($Ms = .07, .10, .43$, and $.52$ for novel, single, medium and large categories for old, compared to $.10, .09, .29$, and $.30$ for young), $F(3, 90) = 5.40$, $MSE = .02$, $p = .002$, for the uncorrected analysis, $F(2, 60) = 3.59$, $MSE = .02$, $p = .03$, for the novel-corrected analysis. Furthermore, although there were no overall age-related differences on the measures of response criteria ($Fs < 1$ and <2.1 for B_D''-novel and B_D''-related, respectively), more focused analyses examining response criteria

---

[4] Given the small effect size for the Age × Encoding Condition interaction on the B_D''-related measure ($f = .15$), the power to detect an interaction was low (28%), both in absolute terms (far short of the 80% level recommended by J. Cohen, 1988) and when compared with that for the more robust effect seen on the B_D''-novel measure ($f = .22$, power = 56%).

Table 2

Measures of Sensitivity and Response Bias, Experiment 2: Encoding Manipulation

Condition	Experimental				Control			
	Old		Young		Old		Young	
	A'	B_D''	A'	B_D''	A'	B_D''	A'	B_D''
Item specific memory (hits compared to novel false alarms)								
Single								
M	.80	.59	.85	.55	.85	.35	.90	.53
SD	.13	.44	.09	.58	.06	.64	.06	.44
Medium								
M	.88	.48	.88	.43	.91	−.03	.91	.39
SD	.07	.53	.07	.49	.06	.61	.06	.57
Large								
M	.91	.43	.88	.41	.92	−.12	.92	.36
SD	.05	.48	.10	.51	.05	.58	.05	.51
Item specific memory (hits compared to related false alarms)								
Single								
M	.74	.48	.83	.45	.77	.06	.86	.16
SD	.17	.58	.08	.47	.09	.61	.08	.51
Medium								
M	.71	−.22	.80	−.08	.77	−.69	.83	−.13
SD	.12	.56	.10	.45	.07	.45	.07	.70
Large								
M	.71	−.52	.80	−.14	.72	−.85	.80	−.36
SD	.12	.39	.10	.46	.11	.22	.08	.62

Note. Category size (number of categorically related items presented): for single condition, 1 item per category; for medium condition, 9 items per category; for large condition, 18 items per category.

in the large category condition and particularly comparing hits to related false alarms (i.e., the condition where older adults were found to be most lenient in the earlier experiments) showed that older adults were still significantly more lenient ($M = -.52$) than younger adults ($M = -.14$) for this comparison, $F(1, 30) = 6.43$, $MSE = .18$, $p = .02$.

Correct recognition (hits) and sensitivity. From Figure 2 and Table 2 it can be seen that the encoding manipulation did not improve either correct recognition or measures of sensitivity. Indeed, correct recognition in the experimental condition ($M = .70$) was significantly lower than in the control condition ($M = .81$), $F(1, 60) = 9.00$, $MSE = .06$, $p = .004$, for uncorrected recognition; $F(1, 60) = 6.40$, $MSE = .08$, $p = .01$, for novel-corrected recognition. The outcomes for the measures of sensitivity showed a similar pattern, with a slight but significant decrement found for A'-novel ($Ms = .87$ and .90 for experimental and control, respectively), $F(1, 60) = 6.44$, $MSE = .01$, $p = .01$, and a trend toward impairment found for A'-related ($Ms = .77$ and .79), $F(1, 60) = 2.22$, $MSE = .02$, $p = .14$. Possible factors contributing to this unexpected finding are considered in the *Discussion.*

Restricting attention to the experimental condition alone, older adults tended to show a more marked decrement in recognition of one-of-a-kind items ($Ms = .51, .71$, and .80 for 1, 9, and 18-item categories, respectively) than did younger adults ($Ms = .68, .76$, and .76, respectively), $F(2, 60) = 2.26$, $MSE = .04$, $p = .11$, for the Age × Category Size interaction. Consistent with our previous findings showing impaired recognition of one-of-a-kind items for

older adults, pairwise comparisons within the older experimental group alone showed a significant decrement in recognition for the single-category items relative to both the medium- and large-category items (smallest $F = 5.35$). In contrast, the recognition performance of younger adults did not reliably vary as a function of category size ($Fs < 1.2$). The measures of sensitivity showed no overall age difference in the experimental group for the A'-novel measure ($F < 1$; $Ms = .86$ and .87 for old and young, respectively) but a significant age-related deficit for the A'-related measure ($Ms = .72$ and .81 for old and young, respectively), $F(1, 30) = 9.68$, $MSE = .02$, $p = .004$. Focused comparisons, contrasting older adults' sensitivity for single-category items with that for the medium- or large-category items pointed to significantly depressed sensitivity for one-of-a-kind items among older adults on the A'-novel measure (smallest $F = 5.74$) but no such selective impairment for younger adults ($Fs < 1.2$). Similar pairwise comparisons for the A'-related measure showed no significant decrements for single-category items for either age group (all $Fs < 1.7$).

Discussion

The results of Experiment 2 indicate that providing explicit encouragement, at the time of initial encoding, to notice and attend to perceptual details of items reduced false recognition and counteracted liberal response bias for both older and younger adults but generally was more beneficial for older adults. That perceptually focused encoding should provide some additional resistance to false recognition is consistent with previous findings with younger adults (Marks, 1991) and related findings in source monitoring situations with older adults (Hashtroudi et al., 1994). It is also consistent with a recent proposal that we made, in the broader context of a *constructive memory framework* (Schacter, Norman, & Koutstaal, 1998). In this framework we have emphasized the necessity, not only for adequate binding of the various features of an episode into a single or unified trace (e.g., what color corresponded with a particular object, or where and when a particular event transpired; Chalfonte & Johnson, 1996; Johnson & Chalfonte, 1994) but also adequate separation of similar episodes so as to permit recognition of particular exemplars or incidents from multiple similar ones. The encoding task may have acted to encourage such pattern separation (McClelland, McNaughton, & O'Reilly, 1995; O'Reilly & McClelland, 1994).

Nonetheless, the perceptually focused encoding did not eliminate all age differences: Some residual difficulties for older adults were still apparent. Although false recognition was, indeed, relatively lower in the experimental than the control condition, within the experimental group alone false recognition among older adults still significantly exceeded that of younger adults. Moreover, although the analyses of response criteria showed more conservative responding in the experimental than in the control condition, and also some evidence that the criteria of older adults were more influenced than those of younger adults, the response criteria of the two age groups still manifestly differed in some critical conditions. In particular, considering only the experimental condition, a focused analysis concentrating on the condition in which age differences in criteria might be expected to be most strongly apparent— that involving large categories and in which hits were compared with related false alarms—showed that older adults were substantially more lenient than their younger counterparts. Persistent age

effects were also observed on the measures of sensitivity: Older adults in the experimental group still showed significantly impaired overall sensitivity when hits were compared with related false alarms and selectively impaired sensitivity for one-of-a-kind items when hits were compared with novel false alarms.

An unexpected finding arose with regard to the measures of correct recognition and also sensitivity as indexed by the A'-novel measure (comparing hits with novel false alarms): For these measures, the performance of the control condition (in which no perceptually specific information was provided during encoding) exceeded that in the experimental condition (in which this information was provided). What could account for this apparently paradoxical outcome? Should not the encoding manipulation have increased rather than decreased sensitivity (e.g., Loftus & Kallman, 1979)?

One possibility is that these reversed differences in sensitivity between the experimental and control groups arose because of the differential salience of the category names for these two groups. Although both the experimental and the control participants were provided the category names before the presentation of the pictures, the further context in which this occurred was (necessarily) different for the two groups. For the experimental group, the categorical information was provided together with distinctive information, and the distinctive information comprised the most relevant information (participants were asked to remember the two specified features and also to determine if they noticed them in the subsequently presented pictures). Thus, it is likely that participants' attentional focus was placed primarily on the distinctive information. In contrast, for the control group, the initial search-and-count filler task that preceded the category labels was irrelevant to the subsequent pictures. Thus, for participants in the control group, attention was probably most focused on the category information. To the extent that this is correct, and categorical information differentially affected the control participants, we may have exacerbated gist-based responding among older adults in the control group. Both the extreme leniency shown by older controls on the B_D''-related measure (average of $-.77$ for the medium and large categories combined)—a leniency even more marked than that shown by the older controls in the previous experiment (average of $-.54$)—and the increase in hit rates relative to controls of Experiment 1 (average hit rates of 89% vs. 81%) are consistent with this account.

Another, more general, possibility focuses on the consequences of shifts in response criteria for veridical recognition. The outcomes for the signal detection analyses clearly indicated that participants in the experimental condition used more stringent recognition criteria than those in the control condition; for example, the overall means for the experimental versus control condition for B_D''-novel were .48 vs. .25; for B_D''-related they were $-.003$ (i.e., essentially unbiased responding) vs. $-.30$. Such differences in criteria might also be accompanied by some reductions in veridical recognition, particularly for older adults for whom a larger proportion of their correct recognition responses would otherwise have been based on general similarity information. Consistent with this are the following: (a) the reductions in veridical recognition in the experimental group were numerically (albeit not significantly) greater for older adults (experimental = .67, control = .83) than for younger adults (experimental = .73, control = .79, or reductions of 16% and 6%, respectively), and (b) the

analyses of sensitivity showed a significant decrement for the experimental condition only when hits were compared with novel false alarms (a condition in which conservative responding might tend to work against the experimental condition) but not when hits were compared to related false alarms (a condition in which conservative responding should work at least as much for, as against, the experimental condition, inasmuch as conservatism would also help to decrease false alarms to related items).

Experiment 3: Combined Encoding and Retrieval Manipulations

The outcomes of Experiments 1 and 2 suggested that either altering the conditions at retrieval (providing greater support for careful scrutiny of items on an item-by-item basis) or altering encoding (encouraging attentive, item-specific perceptual processing of each stimulus) improved the memory performance of older adults, reducing false recognition relative to the respective control conditions and encouraging the use of more stringent response criteria. Given these separate demonstrations of the effectiveness of each of these manipulations, might the combination of the two entirely eliminate age differences in false recognition? To answer this question, Experiment 3 examined the effects of combining the encoding and retrieval manipulations, assessing the recognition performance of older and younger adults in the related pictures paradigm under conditions where participants received both the perceptually focused descriptions during encoding and the inducement to careful retrieval monitoring at test.

Method

Participants. The participants were 16 older adults (*M* age = 67.8 years, range = 63–74 years) and 16 younger adults (*M* age = 19.2 years, range = 18–22 years), recruited and screened according to the same medical and neuropsychological criteria as in the previous experiments. Older adults had significantly more years of formal education (*M* = 16.1, range = 12–20) than did younger adults (*M* = 13.1, range = 12–15), $F(1, 30) = 18.32$, $MSE = 3.93$, $p = .0002$.

Experimental design. The experimental design included a between-subjects variable of age (old, young) and a within-subjects variable of category size. For studied items, category size had three levels: 1, 9, or 18 items presented. For nonstudied items, category size had four levels: 0, 1, 9, or 18 related items presented, with the zero or novel-item condition providing a baseline measure of false alarms.

Procedure. With two exceptions, the experimental stimuli and procedure were identical to those of Experiment 2. First, all participants were presented the descriptions during the study phase: As in the experimental group of Experiment 2, participants read each of the descriptions orally as they were presented, and then, during the presentation of the pictures, pressed the "v" key on the computer keyboard if they noticed the characteristics that were referred to in the description. Second, in the test phase (which, as in Experiment 2, occurred 1 week after the study session), all participants received the retrieval instructions that were previously given to the experimental group of Experiment 1. Thus, participants were informed of the possible similarity of new items to studied items and, rather than making a simple old–new response, were required to classify items as old and identical, new but related, or new and unrelated. As in Experiment 1, participants were asked to verbally state the name of the studied categories to which items that they designated as new but related belonged. In addition, after providing their old and identical, new but related, or new and unrelated judgment, participants were asked to indicate their level of confidence in their judgments.
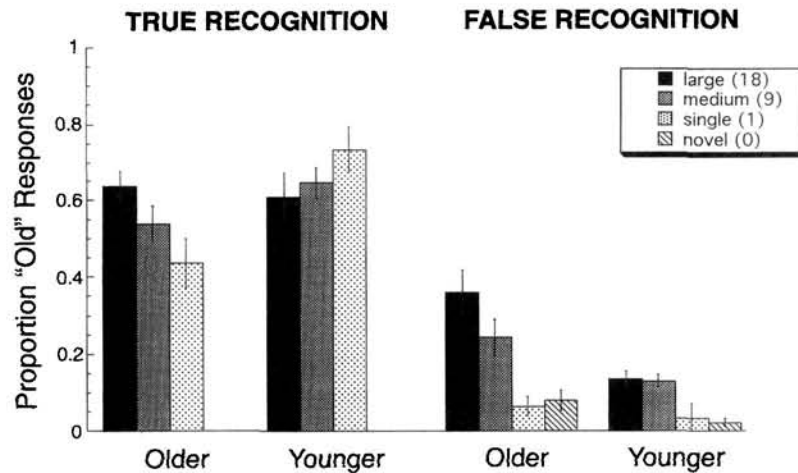
Figure 3. Mean proportion of old responses in Experiment 3 for studied items (true recognition) and nonstudied items (false recognition) as a function of category size and age. Category size, or the number of categorically related items presented during study, was 0, 1, 9, or 18 items; the category size of 0 provides the baseline false alarm rate. All participants were given both the perceptually focused verbal descriptions at encoding and the tri-part retrieval test format. Error bars show the standard error of the mean.

Results

The proportion of correct and false recognition responses are shown in Figure 3. Table 3 gives the outcomes for the signal detection measures.

False recognition and response bias. Encouraging both spe-

Table 3
Measures of Sensitivity and Response Bias, Experiment 3: Encoding and Retrieval Manipulated

	Old		Young	
Condition	A'	B_D''	A'	B_D''
Item specific memory (hits compared to novel false alarms)				
Single				
M	.79	.71	.90	.67
SD	.07	.45	.06	.27
Medium				
M	.83	.68	.88	.79
SD	.05	.41	.05	.17
Large				
M	.85	.62	.86	.75
SD	.07	.39	.09	.26
Item specific memory (hits compared to related false alarms)				
Single				
M	.75	.69	.86	.43
SD	.11	.38	.12	.41
Medium				
M	.72	.35	.83	.45
SD	.11	.54	.08	.31
Large				
M	.70	.03	.80	.43
SD	.13	.61	.12	.53

Note. Category size (number of categorically related items presented): for single condition, 1 item per category; for medium condition, 9 items per category; for large condition, 18 items per category.

cific, perceptually focused encoding and careful monitoring during retrieval resulted in comparatively low rates of false recognition: On average, some 16% of the categorized lures were falsely recognized, relative to 5% of the novel items. (For comparison, the corresponding values in the experimental condition of Experiment 2, involving only the encoding manipulation but otherwise similar conditions, were 29% and 9%, respectively; those for the control condition were 37% and 9%.) Nonetheless, false recognition of older adults still significantly exceeded that of younger adults, both overall (combining across the novel, single, medium, and large categories, $Ms = .19$ and .08), $F(1, 30) = 8.37$, $MSE = .04$, $p = .007$, and following correction for false alarms to novel-category items, $F(1, 30) = 5.98$, $MSE = .02$, $p = .02$. In addition, older adults were more strongly affected by the number of categorically related items that had been studied, showing a considerable increment in false recognition for the 18- compared to 9-item categories (for 0, 1, 9, and 18 items, $Ms = .08, .06, .25$, and .36, respectively), whereas younger adults were less affected ($Ms = .03, .04, .13$, and .14), $F(3, 90) = 5.65$, $MSE = .01$, $p = .001$, and $F(2, 60) = 5.60$, $MSE = .01$, $p = .006$, for the uncorrected and novel-corrected Age × Category Size interactions, respectively. Furthermore, although the analyses of response criteria showed no overall age differences on either the B_D''-novel or the B_D''-related measure ($Fs < 1$), there was a trend toward an Age × Category Size interaction on the B_D''-novel measure, $F(2, 60) = 2.60$, $MSE = .03$, $p = .08$, and a significant Age × Category Size interaction for B_D''-related, $F(2, 60) = 6.64$, $MSE = .13$, $p = .003$. On the B_D''-novel measure, although quite conservative overall, older adults were relatively less stringent for the medium and large categories ($Ms = .68$ and .62) than for the single items ($M = .71$), whereas the reverse was true for younger adults, who were more stringent for items from medium and large (.79 and .75) categories than for single items (.67). More pronounced differences were seen on the B_D''-related measure, with older adults showing substantial effects of category size (means of .69, .35, and .03 for single,

medium, and large, respectively) but younger adults essentially unaffected (.43, .45, and .43, respectively). (Note, however, that older adults were essentially unbiased for the large-category condition.)

Correct recognition (hits) and sensitivity. Analyses of correct recognition revealed an overall age-related deficit in recognition (Ms = .67 and .54 for younger and older adults, respectively), $F(1, 30) = 4.38$, $MSE = .09$, $p = .05$, and $F(1, 30) = 9.75$, $MSE = .08$, $p = .004$, for the uncorrected and novel-corrected responses, respectively. There was also a significant Age × Category Size interaction, $F(2, 60) = 9.08$, $MSE = .02$, $p = .0004$: Older adults showed especially depressed recognition for the one-of-a-kind items (Ms = .44, .54, and .64 for single, medium, and large, respectively), whereas younger adults showed the reverse pattern (higher recognition for single items than for many-exemplar items, Ms = .74, .65, and .61, respectively). The measure of sensitivity comparing hits with novel false alarms showed a similar pattern: There was both an overall age-related deficit in sensitivity, $F(1, 30) = 8.63$, $MSE = .009$, $p = .006$ (for younger and older adults, Ms = .88 and .82, respectively) and an Age × Category Size interaction, $F(2, 60) = 8.20$, $MSE = .002$, $p = .0007$, with the interaction again reflecting older adults' depressed sensitivity for single items (Ms = .79, .83, and .85) but the reverse for younger adults (.90, .88, and .86, respectively). There was also an overall age-related sensitivity deficit on the A'-Related measure, $F(1, 30) = 13.39$, $MSE = .02$, $p = .001$, means for younger and older adults of .83 versus .73; however, for this measure, given the relatively lower levels of false alarms to the single items, both age groups showed somewhat greater sensitivity for the single items (.75, .72, and .70 for older adults, .86, .83, and .80 for younger adults; $F < 1$ for the Age × Category Size interaction).

Discussion

Based on our previous findings showing the beneficial effects of altering retrieval conditions (Experiment 1) or encoding conditions (Experiment 2) when each was altered separately, in this experiment we examined the consequences of their conjoint application. Although, as expected on the basis of the results from Experiments 1 and 2, overall levels of false recognition were low, older individuals still showed significantly elevated false recognition relative to the young, and the old–young discrepancy in the rate of false recognition increased as a consequence of having encountered an increasing number of categorically related exemplars. Whereas false recognition was 1.5 times greater among older adults than younger adults when only one categorically related item had been encountered, older adults made nearly twice as many false recognition responses as the young after exposure to 9 related items, and 2.5 times more when 18 related items had been shown. Moreover, the additional encoding support did not eliminate age differences in overall sensitivity or (considering absolute hits) the tendency to less often correctly recognize one-of-a-kind items than items from multiple-exemplar categories.

It is important to note that a consideration of only those false recognition responses that were accompanied by high confidence (ratings of 4 or 5 on a scale where 1 = *guessing*, 5 = *very sure*) revealed a pattern similar to that observed for the overall responses (see Table 4, which presents the high confidence responses for all three experiments). This suggests that these residual age differ-

ences in false recognition responses—present despite the provision of both encoding and retrieval support—did not reflect only tentatively endorsed items. Rather, in a fair number of instances, the falsely recognized lures among older adults were accepted with high confidence. The implications of these observations are discussed next.

General Discussion

The three experiments reported here demonstrate that *either* providing older adults with support to engage in careful scrutiny and careful decision making regarding items at the time of test, *or* inducing them to adopt a more perceptually focused encoding orientation, may reduce gist-based false recognition. Both retrieval support (Experiment 1) and encoding support (Experiment 2) reduced age-related differences in false recognition, bringing older adults' false recognition rate closer to that observed in younger adults. Yet these manipulations did not suppress the rate of false recognition among older adults to the level shown by the younger age group. Combining the two manipulations also did not reduce older adults' false recognition to a level corresponding to that of the young (Experiment 3). These outcomes clearly imply that susceptibility to false recognition in elderly individuals cannot readily be explained by a simple one-variable account. Rather, it is likely that false recognition reflects multiple and possibly quite subtle variables where alterations of any one component may reduce but not abolish age-related differences.

Some of the persistent age differences we observed in false recognition may be attributable to residual age differences in response criteria. Although the retrieval manipulation clearly and unambiguously reduced overall biases toward saying "old" and often differentially helped elderly adults, age differences in response criteria were still sometimes present, particularly for the large (18-exemplar) categories. Older and younger adults often showed similar or comparatively similar criteria for single items (cf. Isingrini et al., 1995; Till, Bartlett, & Doyle, 1982). In the case of multiple exemplars, it is likely that both older and younger adults are aware that multiple exemplars were encountered previously. In the absence of countervailing instructions, both old and young respond somewhat leniently in this situation. With encouragement, at the time of retrieval, to use *yes* responses cautiously, both age groups may shift their criteria, but older adults may remain less able to resist or oppose familiarity arising from general category information or from other variables, including perceptual similarity information, than younger adults (cf. Bartlett, Strater, & Fulton, 1991).

It is also possible that residual age differences in false recognition can be attributed to differences in the quality of memory in older versus younger adults. Within the context of these experiments, we quite consistently found age-related deficits in sensitivity for the measures of item-specific memory, and age deficits were also often found on correct recognition and high confidence correct recognition (Table 4). More generally, there is evidence that older adults may less often specifically recollect or "remember" items than do younger adults (Mäntylä, 1993; Parkin & Walter, 1992; Perfect, Williams, & Anderton-Brown, 1995), and specific probes of the qualitative characteristics of their memory suggest that they may have access to fewer perceptual details regarding study items (e.g., Hashtroudi, Johnson, & Chrosniak,

Table 4

High Confidence Hits and False Alarms, Experiments 1–3

Experiment & condition	Hits						Within category false alarms						Novel category false alarms	
	Single		Medium		Large		Single		Medium		Large		Zero	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Experiment 1: Retrieval														
Experimental														
Old	.41	.25	.55	.23	.56	.23	.08	.12	.26	.18	.26	.19	.05	.09
Young	.58	.25	.56	.24	.52	.21	.00	.00	.08	.11	.10	.08	.004	.02
Control														
Old	.46	.25	.60	.22	.63	.22	.18	.16	.35	.20	.46	.25	.11	.13
Young	.73	.21	.62	.17	.63	.17	.04	.11	.09	.11	.09	.12	.02	.03
Experiment 2: Encoding														
Experimental														
Old	.36	.30	.53	.24	.65	.20	.06	.12	.25	.20	.35	.15	.03	.05
Young	.51	.24	.58	.13	.55	.16	.04	.08	.15	.10	.15	.11	.03	.07
Control														
Old	.53	.29	.75	.18	.76	.18	.15	.17	.39	.22	.54	.25	.11	.16
Young	.48	.28	.53	.24	.60	.25	.04	.08	.13	.10	.19	.14	.004	.02
Experiment 3: Encoding and Retrieval														
Old	.34	.19	.45	.20	.56	.18	.05	.09	.21	.19	.28	.21	.04	.09
Young	.56	.23	.51	.19	.48	.26	.01	.05	.06	.08	.06	.05	.004	.02

Note. Category size (number of categorically related items presented) is as follows: For zero condition, 0 items per category; for single condition, 1 item per category; for medium condition, 9 items per category; for large condition, 18 items per category. Means are for responses accompanied by a confidence rating of 4 or 5 on a scale where 1 = *guessing*, 5 = *very sure*. Experiment 1 control data are from Koutstaal and Schacter (1997, Experiment 3).

1990; Hashtroudi et al., 1994; Koutstaal, Schacter, Johnson, Angell, & Gross, 1998; Norman & Schacter, 1997).

Such differences in the quality of memory representations available to older adults might contribute to differences in false recognition in any number of ways. One possibility, well supported by internal and external considerations, relates to the effects of item-specific memory on response criteria. If younger adults have more, or more distinctive, recollective material available to them for some or most studied items, they can use a more stringent criterion, demanding more of an item before they designate it as old (Norman & Schacter, 1997; Schacter, Israel, & Racine, 1999; Schacter, Norman, & Koutstaal, 1998; also cf. Loftus & Bell, 1975). It is possible that the differences in response criteria for younger versus older adults in the control conditions partially derive from this source: For younger adults, the clearly remembered details of some of the studied items may act to reinforce the notion that "much can be demanded" and may counteract a tendency to drift toward using more lenient familiarity-based criteria. In addition, it is possible that a similar and stronger form of this effect may have been induced among both older and younger adults through the verbal descriptions intended to promote more item-specific perceptual processing of the items.

Consistent with the foregoing considerations, an across-experiment comparison of participants' response criteria on the B_D''-related measure, comparing hits with related false alarms for the medium and large categories combined, showed that although participants were most stringent when retrieval support was provided, the encoding manipulation also clearly added something: For the encoding manipulation alone, older adults used more stringent criteria than for either of the control conditions (encoding or retrieval controls) and, for both age groups, criteria were more stringent when both encoding and retrieval support were present than with retrieval support alone (for control-encoding, control-retrieval, experimental-encoding, experimental-retrieval, and experimental-encoding-and-retrieval, $Ms = -.77, -.54, -.37, .08,$ and $.19$, respectively, for older adults; corresponding $Ms = -.25, -.14, -.11, .22,$ and $.44$, respectively, for younger adults).

The possibility that both older and younger adults may sometimes respond to manipulations designed to enhance item-specific encoding with a quite general or global shift in criteria has also been raised by recent work in a somewhat different paradigm. Schacter et al. (1999; Israel & Schacter, 1997) compared the false recognition responses of older and younger adults in the Deese–Roediger-McDermott verbal converging associates paradigm under conditions in which the list items were encoded with little distinctive information (the words-only condition, in which participants were presented words visually and auditorily), with a situation where the list items were accompanied by distinctive pictorial information (the words + pictures condition, in which participants saw a black-and-white line drawing corresponding to

the word and heard the word presented auditorily). Under these conditions, both older and younger adults showed a significant reduction in false recognition after pictorial encoding compared with word encoding. Critically, however, this reduction in false recognition was observed only when the pictorial manipulation was conducted on a between-groups basis. When, instead, a given participant was presented some study lists with accompanying pictures and others without such distinctive pictorial information, then neither older nor younger adults showed a reduction of false recognition after picture compared with word encoding. Schacter et al. (1999) hypothesized that this finding may have arisen as a consequence of an overall or global shift in response criteria. Participants in the between-groups paradigm may have relied on what they termed a *distinctiveness heuristic:* a mode of responding in which, because they are aware that they can sometimes recollect items with a high degree of specificity, participants consistently demand detailed, high-quality recollection to justify a positive recognition response. To the extent that the verbal descriptions in Experiments 2 and 3 acted in a similar manner, both the reductions in false recognition and the relative reductions in correct recognition and sensitivity seen in the experimental condition might follow.

In our previous experiments (Koutstaal & Schacter, 1997), we also computed a measure of sensitivity to gist information, treating false alarms to categorically related items as evidence of a form of generic memory and comparing such *gist memory responses* with false alarms to novel-category items. This measure, however, does not provide a measure of gist memory per se, as participants might be aware that a new item is categorically similar to items that were shown at study but correctly reject it on the grounds that they had no specific recollection of that particular item. In other words, false alarms to related items provide a measure of willingness to respond on the basis of gist information, given that such responding was not opposed by item-specific information. To the extent that younger adults generally have more detailed recollection of the study items (i.e., greater item-specific memory), such opposition would occur more frequently for younger than for older adults, possibly inflating differences in the estimate of gist memory for older compared with younger adults.

However, the tri-part retrieval classification task used in Experiments 1 and 3 provides a unique opportunity to address the question of the amount of gist memory possessed by participants versus their willingness to respond on the basis of gist. In particular, an estimate of memory for gist (rather than unopposed willingness to respond to gist) might be obtained by combining responses to four types of items: *hits* (which might be based on either item-specific or gist information, but where access to the former would also generally allow access to the latter), *related false alarms*, correct rejections where participants indicated that the lure was new but related to the presented items (*new but related correct rejections*), and *related misses* (studied items designated as new but related). Combining across all of these response categories, but excluding single items (for which gist-related responding was probably minimal), we found no age-related differences in gist memory in either experiment (for old and young experimental participants in Experiment 1, *M*s = .97 and .98, respectively; in Experiment 3, *M*s = .91 and .93, respectively; means are based on the sum of the four response categories divided by 2, as both studied and nonstudied items could contribute to the averages).

Although the lack of age differences on this measure might reflect ceiling effects, we also found clear evidence of age differences in *willingness to respond affirmatively* on the basis of such gist information. Specifically, the proportion of related false alarms out of all responses that may have involved gist memory provides a measure of gist-based responding that takes into account any initial differences in memory for gist. This measure [(old responses to related new items)/½ (old responses to categorized studied items + old responses to related new items + correct rejections of new but related items + misses categorized as new but related)] showed significantly greater gist-based responding by older than by younger adults for both of the experiments in which the tri-part retrieval classification was included: Experiment 1, for the experimental group alone, average proportions for old and young, *M*s = .33 and .18, respectively, $F(1, 30) = 8.90$, $MSE = .04$, $p = .006$; Experiment 3, average proportions for old and young, *M*s = .33 and .14, respectively, $F(1, 30) = 10.74$, $MSE = .05$, $p = .003$.

These analyses underscore the residual age-related differences in false recognition that remained, indicating that—despite altered conditions at retrieval or at both encoding and retrieval—older adults remained more willing than were younger adults to designate items as having been previously encountered when they shared general similarity with items they had seen. However, in both of these experiments, the comparisons with the control conditions showed that substantial reductions in false recognition did occur. Our results thus both clearly demonstrate elevated similarity- or gist-based false recognition among older adults and show that there are concrete and practically realizable steps that can be taken to reduce such errors. In this, they converge with several other recent efforts to determine the conditions under which false recognition and false recall may be minimized or rendered less likely to occur, both in individuals with intact memory functioning (e.g., Gallo, Roberts, & Seamon, 1997; McDermott, 1996; Read, 1996; Tussing & Greene, 1997; also see Gauld & Stephenson, 1967) and in individuals with amnesia (Schacter, Verfaellie, Anes, & Racine, 1998) and conditions involving disruptions of frontal and inhibitory functions (Curran et al., 1997; Parkin, Bindschaedler, Harsent, & Metzler, 1996; Schacter, Curran, Galluccio, Milberg, & Bates, 1996). Such efforts should yield further data and testing grounds for understanding constructive aspects of memory (Johnson et al., 1993; Reyna & Brainerd, 1995; Schacter, Norman, & Koutstaal, 1998) which, in turn, may allow an increasingly articulated characterization of human memory to emerge—one ignoring neither its susceptibility to error nor its general faithfulness, comprising, as it still does, a valued and often reliable guide to judgment and action.

References

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A' and B" to below-chance performance. *Psychological Bulletin, 102,* 439–442.

Bartlett, J. C., Leslie, J. E., Tubbs, A., & Fulton, A. (1989). Aging and memory for pictures of faces. *Psychology and Aging, 4,* 276–283.

Bartlett, J. C., Strater, L., & Fulton, A. (1991). False recency and false frame of faces in young adulthood and old age. *Memory & Cognition, 19,* 177–188.

Bower, G. H., & Glass, A. L. (1976). Structural units and the redintegrative power of picture fragments. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 456–466.

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34,* 157–185.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24,* 403–416.

Cohen, G., & Faulkner, D. (1989). Age differences in source forgetting: Effects on reality monitoring and eyewitness testimony. *Psychology and Aging, 4,* 10–17.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & Rabinowitz, J. C. (1985). Age differences in the acquisition and use of verbal information: A tutorial review. In H. Bouma and D. G. Bouwhuis (Eds.), *Attention and performance* (Vol. 10, pp. 471–499). Hillsdale, NJ: Erlbaum.

Curran, T., Schacter, D. L., Norman, K. A., & Galluccio, L. (1997). False recognition after a right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia, 35,* 1035–1049.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Dewhurst, S. A., & Conway, M. A. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1088–1098.

Dodson, C. S., & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology, 106,* 541–557.

Donaldson, W. (1993). Accuracy of d' and A' as estimates of sensitivity. *Bulletin of the Psychonomic Society, 31,* 271–274.

Dywan, J., & Jacoby, L. (1990). Effects of aging on source monitoring: Differences in susceptibility to false fame. *Psychology and Aging, 5,* 379–387.

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 and Memory, 6,* 588–598.

Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review, 104,* 148–169.

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.

Gauld, A., & Stephenson, G. M. (1967). Some experiments relating to Bartlett's theory of remembering. *British Journal of Psychology, 58,* 39–49.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424–429.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging, 5,* 119–126.

Hashtroudi, S., Johnson, M. K., Vnek, N., & Ferguson, S. A. (1994). Aging and the effects of affective and factual focus on source monitoring and recall. *Psychology and Aging, 9,* 160–170.

Henkel, L. A., Johnson, M. K., & DeLeonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127,* 251–268.

Hess, T. M. (1984). Effects of semantically related and unrelated contexts on recognition memory of different-aged adults. *Journal of Gerontology, 39,* 444–451.

Hess, T. M. (1990). Aging and schematic influences on memory. In T. M. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 93–160). Amsterdam, The Netherlands: Elsevier.

Hess, T. M., & Higgins, J. N. (1983). Context utilization in young and old adults. *Journal of Gerontology, 38,* 65–71.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory

in a multiple-trace theory of memory. *Psychological Review, 95,* 528–551.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33,* 1–18.

Hintzman, D. L., & Curran, T. (1995). When encoding fails: Instructions, feedback, and registration without learning. *Memory & Cognition, 23,* 213–226.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74,* 351–354.

Isingrini, M., Fontaine, R., Taconnat, L., & Duportal, A. (1995). Aging and encoding in memory: False alarms and decision criteria in a word-pair recognition task. *International Journal of Aging and Human Development, 41,* 79–88.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4,* 577–581.

Johnson, M. K., & Chalfonte, B. L. (1994). Binding complex memories: The role of reactivation and the hippocampus. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 311–350). Cambridge, MA: MIT Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37,* 555–583.

Koutstaal, W., Schacter, D. L., Johnson, M. K., Angell, K. E., & Gross, M. S. (1998). Post-event review in older and younger adults: Improving memory accessibility of complex everyday events. *Psychology and Aging, 13,* 277–296.

Labouvie-Vief, G. (1980). Adaptive dimensions of adult cognition. In N. Datan & N. Lohmann (Eds.), *Transitions of aging* (pp. 3–26). New York: Academic Press.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17,* 349–358.

Loftus, G. R., & Bell, S. M. (1975). Two types of information in picture memory. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 103–113.

Loftus, G. R., & Kallman, H. J. (1979). Encoding and use of detail information in picture recognition. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 197–211.

Mäntylä, T. (1993). Knowing but not remembering: Adult age differences in recollective experience. *Memory & Cognition, 21,* 379–388.

Marks, W. (1991). Effects of encoding the perceptual features of pictures on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 566–577.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition, 25,* 826–837.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35,* 212–230.

Multhaup, K. S. (1995). Aging, source, and decision criteria: When false fame errors do and do not occur. *Psychology and Aging, 10,* 492–497.

Naveh-Benjamin, M., & Craik, F. I. M. (1995). Memory for context and its use in item memory: Comparisons of younger and older persons. *Psychology and Aging, 10,* 284–293.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and

older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25,* 838–848.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus, 4,* 661–682.

Park, D. C., Puglisi, J. T., & Smith, A. D. (1986). Memory for pictures: Does an age-related decline exist? *Journal of Psychology and Aging, 1,* 11–17.

Parkin, A. J., Bindschaedler, C., Harsent, L., & Metzler, C. (1996). Verification impairment in the generation of memory deficit following ruptured aneurysm of the anterior communicating artery. *Brain and Cognition, 32,* 14–27.

Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging, 7,* 290–298.

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.

Perfect, T. J., Williams, R. B., & Anderton-Brown, C. (1995). Age differences in reported recollective experience are due to encoding effects, not response bias. *Memory, 3,* 169–186.

Rabinowitz, J. C., & Ackerman, B. P. (1982). General encoding of episodic events by elderly adults. In F. I. M. Craik & S. Trehub, *Aging and cognitive processes* (pp. 145–154). New York: Plenum Press.

Rabinowitz, J. C., Craik, F. I. M., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology, 36,* 325–344.

Rankin, J. L., & Kausler, D. H. (1979). Adult age differences in false recognitions. *Journal of Gerontology, 34,* 58–65.

Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review, 3,* 105–111.

Reder, L. M., Wible, C., & Martin, J. (1986). Differential memory changes with age: Exact retrieval versus plausible inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 72–81.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim thesis. *Learning and Individual Differences, 7,* 1–75.

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. (1996). Memory illusions. *Journal of Memory and Language, 35,* 76–100.

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.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Schacter, D. L. (1995). Memory distortion: History and current status. In D. L. Schacter, J. T. Coyle, G. D. Fischbach, M. M. Mesulam, & L. E. Sullivan (Eds.), *Memory distortion: How minds, brains, and societies reconstruct the past* (pp. 1–43). Cambridge, MA: Harvard University Press.

Schacter, D. L., Curran, T., Galluccio, L., Milberg, W., & Bates, J. (1996). False recognition and the frontal lobe: A case study. *Neuropsychologia, 34,* 793–808.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recog-

nition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40,* 1–24.

Schacter, D. L., Koutstaal, W., Johnson, M. K., Gross, M. S., & Angell, K. E. (1997). False recollection induced by photographs: A comparison of older and younger adults. *Psychology and Aging, 12,* 203–215.

Schacter, D. L., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences, 1,* 229–236.

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., Verfaellie, M., Anes, M. D., & Racine, C. (1998). When true recognition suppresses false recognition: Evidence from amnesic patients. *Journal of Cognitive Neuroscience, 10,* 668–679.

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.

Schramke, C. J., & Bauer, R. M. (1997). State-dependent learning in older and younger adults. *Psychology and Aging, 12,* 255–262.

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 267–287.

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

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10,* 527–539.

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language, 33,* 203–217.

Till, R. E., Bartlett, J. C., & Doyle, A. H. (1982). Age differences in picture memory with resemblance and discrimination tasks. *Experimental Aging Research, 8,* 179–184.

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

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13,* 230–241.

Tussing, A. A., & Greene, R. L. (1997). False recognition of associates: How robust is the effect? *Psychonomic Bulletin & Review, 4,* 572–576.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70,* 122–129.

Weldon, M. S., & Roediger, H. L., III. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition, 15,* 269–280.

Zaragoza, M. S., & Koshmider, J. W., III. (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 246–255.