JOURNAL OF MEMORY AND LANGUAGE **37,** 555–583 (1997)
ARTICLE NO. ML972529

Gist-Based False Recognition of Pictures in Older and Younger Adults

Wilma Koutstaal and Daniel L. Schacter

Harvard University

Three experiments examined the recognition performance of older (60–75 yrs) and younger (17–25 yrs) adults for detailed colored pictures of objects as a function of whether the targets and lures were from previously studied categories or were unrelated (noncategorized) items. If participants had studied many exemplars from a category, they often falsely recognized lures from those categories; this false recognition effect was especially pronounced in older adults. Older and younger participants showed equivalent correct recognition of large-category targets, but older adults showed significantly reduced recognition of unrelated targets, suggesting that older adults were relying on the general conceptual and/or perceptual similarity (''gist'') of study and test items in making their recognition decisions. The results extend the domain of robust false recognition to detailed color pictures and demonstrate that, particularly in older adults, false recognition sometimes involves similarity-based errors rather than source confusions regarding whether specific lure items had been presented or were generated spontaneously during the study task.

False recognition is typically observed when novel items are in some way related to previously experienced items. A ''new'' or not-previously-encountered statement, for example, may be mistakenly designated as ''old'' because it is highly consistent with statements that were, in fact, presented or because it was a pragmatic or logical inference that could be drawn from those statements (e.g., Bransford & Franks, 1971; Johnson, Bransford, & Solomon, 1973; Sulin & Dooling, 1974). Similarly, a word might be erroneously designated as ''old'' because it is associatively or categorically related to one or more previously presented items (Anisfeld & Knapp, 1968; Paul, 1979; Shiffrin, Huber, & Marinelli, 1995; Underwood, 1965; Vogt & Kimble, 1973). Items may be correctly—or incorrectly—judged as previously encountered because they are broadly consistent with the conceptual or perceptual features of things that were studied, largely matching the overall themes or predominant categories of earlier encountered words, statements, or objects (for review, see Roediger, 1996; Schacter, 1995; Schacter, Norman, & Koutstaal, in press).

Using a method of multiple ''converging associates'' originally developed by Deese (1959) to explore intrusions during free recall, recent research has shown the emergence of high rates of false recognition in the verbal domain (Roediger & McDermott, 1995). In the initial variant of this paradigm, individuals were asked to listen to several sets of words. Each set was constructed around a single theme, such that participants heard numerous words, each of which, on the basis of independent word association norms, was determined to be related to the theme (e.g., ''thread,'' ''pin,'' ''eye,'' ''sew,'' and ''sharp'' are all associated to the word ''needle''). However, the one word upon which all of these presented items associatively converge is not presented: thus the word ''needle'' is never read to the participants. Later, participants are very likely to incorrectly designate these ''critical lures'' as having been presented on the study list; they are also quite likely to produce these items as intrusions during free recall. For ex-

This research was supported by National Institute on Aging Grant AG08441. We thank David Ekstrom, Lissa Galluccio, and Mara Gross for help in data collection.

Address correspondence and reprint requests to Wilma Koutstaal, Department of Psychology, William James Hall, 33 Kirkland Street, Cambridge, MA 02138. Fax: (617) 496-3122; E-mail: wk@wjh.harvard.edu.

0749-596X/97 $25.00

ample, rates of false recognition as high as 84% have been reported (Roediger & McDermott, 1995), with false recognition rates often approaching the level of true recognition (Mather, Henkel, & Johnson, in press; Norman & Schacter, in press; Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996).

Might false recognition also be observed using *pictorial* rather than verbal stimuli? The remarkable capacity that individuals possess for accurate picture recognition has long been known, with excellent retention reported for sets of 2500 and even 10,000 pictures (Shepard, 1967; Standing, Conezio, & Haber, 1970; Standing, 1973). Yet the extent to which false recognition of pictures may be induced is less clear. The pictorial stimuli used by Shepard (1967) were selected so as to be of collectively low similarity and confusability; this was not, however, true for the stimuli employed by Standing et al. (1970), who specifically remarked on the low levels of interference found for certain types of very similar pictures (e.g., some 300 of 2560 pictures represented pictures of single adult male figures; also see Standing, 1973, Experiment 2). However, levels of accurate recognition were considerably reduced when numerous stimuli from a single class, such as ink blots or snow crystals, were employed (Goldstein & Chance, 1970; cf. Wyant, Banks, Berger, & Wright, 1972). Bower and Glass (1976) found that in a two-alternative forced-choice recognition test for line drawings, subjects made significantly more errors and were significantly less confident on test pairs with structurally similar rather than structurally dissimilar distractors. More recently, Strack and Bless (1994, Experiment 1) reported somewhat more strongly elevated rates of false recognition when individuals had been presented numerous black-and-white slides depicting objects drawn from a single class of objects (tools) and were given a verbal recognition test including lures from the same large class. Whereas false recognition responses to items that were not drawn from the studied class (e.g., a shoe, a book) were virtually nonexistent, the average level of false recognition of within-class lures was 28%.

The first of two key questions we asked in the present experiments concerns whether young adults could be led to make high levels of false-recognition errors when highly detailed pictures serve as the stimuli. The experiments reported here address this question through the use of a new picture recognition paradigm. Individuals were shown richly colored and highly detailed pictures of various objects, animals, and people, such as chairs, cats, whales, men, children, and so on. Each picture represented a single object (or, in a few cases, a coherent grouping of objects) with no background setting or scene presented (cf., for example, Frieske & Park, 1993; Intraub, Bender, & Mangels, 1992; Park & Whitten, 1977; Pezdek, Maki, Valencia-Laver, Whetstone, Stoeckert, & Dougherty, 1988; Phelps & Gazzaniga, 1992). For some object categories, relatively few exemplars of the category were shown; for other categories, a comparatively greater number of items was shown. Also interspersed among these items were unrelated items, that is, items that did not belong to any of the categories (e.g., a snow sled, a wine jar). Later, individuals were shown a subset of the studied items, together with new items that either belonged to the studied categories (within-category lures) or that were from novel, nonstudied categories (novel-category lures); they were also presented isolated or noncategorized unrelated items, with some of these items having been studied earlier (unrelated targets) and some presented only at test (unrelated lures). Here, one focus of interest concerns the rates of false recognition for within-category lures—that is, new distractor items that are drawn from the same categories as studied items. A further focus of interest, however, concerns *correct recognition* performance for the unrelated items relative to categorized items. If participants were employing their knowledge of the general classes of things that they had studied to support their positive recognition decisions, then one would anticipate relatively higher proportions of ‘‘old’’ responses for *target* items that were drawn from studied categories than for unrelated (noncategorized) items (cf. Hintzman, 1988).

The second key question we address concerns the performance of younger versus older adults in our paradigm. Compared to younger adults, older adults sometimes show a heightened propensity to falsely recall and recognize the ‘‘critical lures’’ in the paradigm originally developed by Deese (1959) and recently extended by Roediger and McDermott (1995). Relative to their rate of correct recognition, older adults are in some instances more likely to incorrectly designate as ‘‘old’’ the single item on which many target items converge but that is not itself presented (Norman and Schacter, in press; Tun, Wingfield, Rosen, & Blanchard, 1996) and are also more likely to falsely recognize items that were comparatively more weakly associated with the target items (Tun et al., 1996). Older adults have also been shown to be more prone to falsely recognize synonyms or semantic associates of list items than are young adults (Rankin & Kausler, 1979; Smith, 1975; for recent review see Schacter, Koutstaal, & Norman, 1997).

Early interpretations of the latter sets of findings focused on how stimuli are encoded by elderly adults. Compared to their younger counterparts, older adults were hypothesized to process stimuli in a less specific or more generic fashion, noticing and attending to fewer distinctive perceptual or conceptual features of each stimulus (Rabinowitz & Ackerman, 1982; Rabinowitz, Craik, & Ackerman, 1982; see also Hess, 1984; Isingrini, Fontaine, Taconnat, & Duportal, 1995). More recent accounts of elevated susceptibility to inaccuracies of recognition in the elderly have, however, tended to emphasize other factors in accounting for their performance. Deficits in source monitoring (Johnson, Hashtroudi, & Lindsay, 1993) or the ability to assess and determine the origins of past experiences have been especially emphasized. Older adults have been found, under certain conditions, to experience greater difficulties than the young in correctly ascribing past experiences to specific sources or situations, such as determining which person read a particular fact (e.g., Schacter, Kaszniak, Kihlstrom, & Valdiserri, 1991; McIntyre & Craik, 1987), indicating whether they had said or imagined saying a particular word (Hashtroudi, Johnson, & Chrosniak, 1989), or determining whether they had imagined performing a given action or had watched someone else do so (e.g., Cohen & Faulkner, 1989; for recent review, see Spencer & Raz, 1995). Such source monitoring difficulties may clearly contribute to several of the documented instances of an age-related elevation in false recognition, including enhanced susceptibility to errors in the misleading postevent information paradigm (Cohen & Faulkner, 1989). A recently reported finding from our laboratory that older adults are more likely than younger adults to falsely ‘‘remember’’ having watched an event on a videotape when, in fact, they only viewed a photograph of the event (Schacter, Koutstaal, Johnson, Gross, & Angell, 1997) is also readily incorporated into a source monitoring deficit account. In both the Cohen and Faulkner (1989) experiment and the Schacter et al. (1997) study, elderly adults confused the source of a verbal suggestion or picture that they had, in fact, experienced previously.

Source confusions of this kind might also contribute to false recognition in the Deese/Roediger–McDermott paradigm, described earlier. One account of the high levels of recall intrusions and false recognition produced by this paradigm—an account originally emphasized by Deese (1959) and others (Bousfield, Whitmarsh, & Danick, 1958; Underwood, 1965)—focused on the notion of ‘‘implicit associative responses.’’ Although the critical lure words in this paradigm (e.g., ‘‘needle’’ and ‘‘mountain’’) are not actually presented, participants themselves may spontaneously generate the items during the study phase. That is, even though they are not themselves presented, these critical theme words may ‘‘come to mind’’ during the study phase. To the extent that the later recognition or recall tests require differentiation of items that were actually presented from items that may only have been thought about during the study phase, this poses a difficult source monitoring problem: the individual must remember whether the item was, indeed, actually presented during the study phase or, instead, was possibly ‘‘thought of’’ (internally generated) but never externally presented (cf. Read, 1996; Roediger & McDermott, 1995).

The combination of an age-related enhanced susceptibility to false recognition of semantic associates (Norman & Schacter, in press; Rankin & Kausler, 1979; Smith, 1975; Tun et al., 1996) and the well-documented presence of source confusions in the elderly raises an important question. To what extent would elderly show enhanced susceptibility to false recognition in a paradigm in which it is unlikely that they would specifically generate a particular item at study?

The three experiments reported here addressed this question by comparing the performance of older and younger adults in the picture recognition paradigm. In this paradigm, the ‘‘lures’’ are also highly distinctive and detailed colored pictures and it is exceedingly unlikely that the participants themselves would have ‘‘generated’’ the pictures during the study phase of the experiment. Rather, false recognition of lures from studied categories is likely driven by perceptual and/or conceptual correspondence between the lures and common properties of studied exemplars—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). For example, during the study phase, individuals might notice that several different examples of ‘‘chairs’’ or ‘‘dogs’’ had been presented and might even—at the time of first encountering each picture—mentally classify the pictures according to the categories to which they belong, thinking to themselves, ‘‘dog,’’ ‘‘chair,’’ ‘‘sled,’’ ‘‘whale,’’ etc. Then, at test, they might inappropriately extend this generic conceptual classification to new (not previously encountered) items. Alternatively, or in addition, the perceptual similarities of different items within a given category might contribute to later difficulties in discriminating previously encountered items from similar, but never-presented, items. For example, after encountering numerous exemplars of dogs, the specific features of any new (not previously presented) dog are increasingly likely to resemble features that had been encountered in earlier-presented dogs. Such similarities of features might occur either singly (e.g., a particular within-category lure from the category of dogs might have long ears similar to those of a previously presented dog) or in combination, such that some *conjunctions* of features may have occurred previously (e.g., another within-category lure from the category of dogs has both reddish-brown fur and a long thin nose similar to a previously presented dog). Note that, in a broad or extended sense, errors arising from these forms of conceptual or perceptual similarity can be construed as forms of source monitoring errors, but only in the sense that all false recognition involves ‘‘misattribution’’ of events or stimuli to one’s past on the grounds of *some* conceptual or perceptual similarity. However, unlike other paradigms involving ‘‘source monitoring errors’’ in the strict sense, in our picture-recognition paradigm, the *specific token* presented at test was never previously presented nor implicitly generated by the participant. The particular item presented at test for a recognition decision is, *in fact*, *new* (unlike, for example, an event or word that was itself previously encountered or generated but not in the required context), though it may be misconstrued as ‘‘old’’ if insufficiently specific perceptual or conceptual features are employed in reaching a new/old judgment.

Do elderly adults, compared to younger adults, also show enhanced susceptibility to false recognition under these conditions? If elevated false recognition in elderly individuals is driven, at least in part, by overly general or generic encoding of the target information, then age-related increases in gist-based false recognition should be observed in our paradigm. Such an outcome would also indicate that a source confusion account in the strict sense noted above cannot provide an adequate explanation of all age-related increases in susceptibility to false recognition.

EXPERIMENT 1

The primary aims of this experiment were to determine if susceptibility to false recognition could be demonstrated in the picture recognition paradigm just described and to compare the magnitude of the gist- or similarity-based false recognition effect shown by older, compared to younger, adults. We particularly wanted to examine age differences in susceptibility to false recognition in this paradigm under conditions where the elderly might have a reasonable opportunity to form distinctive records of the studied items. Requiring participants to provide a specific judgment concerning each of the presented stimuli during their initial encounter with the items may help to diminish nonspecific differences in memory performance of older and younger adults that are observed during general but otherwise unstructured intentional encoding instructions (Craik, 1977; Craik, Anderson, Kerr, & Li, 1995; Craik & Rabinowitz, 1985), and may reduce the likelihood of attentional lapses in all participants (cf. Shiffrin et al., 1995). Therefore, in contrast to previously reported experiments with the elderly under the Deese/Roediger–McDermott paradigm, in which no encoding or orienting task is used, we used a ‘‘liking’’ encoding task. Individuals were shown each picture for two seconds and then were asked to rate how much they liked the picture.

Method

Participants. Participants were 16 older (*M* age = 68.7 yrs, range = 63–75) and 16 younger (*M* age = 18.8 yrs, range = 17–21) adults. Older participants were recruited via posters and flyers and were individually interviewed so as to exclude those with any of the following conditions: a history of alcoholism or substance abuse; cerebrovascular accident; recent myocardial infarction; present or previous treatment for psychiatric illness; current treatment with psychoactive medication; metabolic or drug toxicity; primary degenerative brain disorders (e.g., Alzheimer’s disease, Parkinson’s disease, or Huntington’s disease); and brain damage sustained earlier from a known cause (e.g., hypoxia). Younger participants were recruited through sign-up sheets posted at Harvard University and included both Harvard undergraduates and Harvard Summer School students. Older participants had, on average, 15.4 years of formal education (range = 12–20 yrs); younger participants had on average 12.4 years (range = 12–14 yrs). All participants (with the exception of one elderly individual) had normal or corrected-to-normal vision and were paid for their involvement in the study.

Stimuli. The stimuli were colored pictures selected from a series of children’s books containing removable and self-adhesive pictures or ‘‘stickers.’’ The pictures, which varied somewhat in size (smallest of approximately 5 × 5 cm, largest of approximately 11 × 9 cm, with a modal size of approximately 8 × 5 cm), were rich in both level of detail and color, and comprised either single objects or small groups of objects without background. Each picture was mounted in the center of a 12.7 × 20.3 cm plain white index card.

The pictures were drawn from various object categories (e.g., boats, cats, shoes, and teddy bears), and were assigned to one of three category sizes: *small* (comprised of a total of 9 exemplars, with 6 exemplars presented at study and 3 exemplars used as nonstudied within-category lures on the recognition test), *medium* (comprised of a total of 12 exemplars, with 9 exemplars presented at study and 3 exemplars used as nonstudied within-category lures during recognition testing), and *large* (comprised of a total of 21 exemplars, with 18 exemplars presented at study and 3 exemplars used as nonstudied within-category lures during recognition testing). The specific categories used and whether they comprised small, medium, or large categories were determined by the availability of the stimuli, such that some stimulus items (e.g., boats) always constituted ‘‘small’’ categories, whereas others always comprised either ‘‘medium’’ (e.g., shoes) or ‘‘large’’ (e.g., cats) categories. A total of 19 different stimulus categories was formed: 6 small, 5 medium, and 8 large categories. These categories were then counterbalanced through studied or nonstudied status separately for each category size (small, medium, and large), such that separate estimates of correct recognition, related false alarms, and also *novel-category* false alarms were obtained for categories of each size. (More specifically, this procedure was followed for 16 of the 19 categories, including 4 small, 4 medium, and 8 large categories; a subset of stimulus items from one of the small categories and from two of the medium categories was always presented, and these three categories never contributed to the ‘‘novel’’ baseline estimation of false alarms.) In addition, a set of 30 *unrelated* items, comprised of pictures that did not clearly belong to any of the categories (e.g., a hand bell, a snow sled, a painted wine jar, and a South American gold figure), was included. One half of these unrelated items were presented at study whereas the remaining half served as nonstudied unrelated lure items; these items (Set A or Set B) were also separately counterbalanced across studied and nonstudied status.

In order to avoid confounding the effects of the number of items in a category that were present at *study,* and the number of items that were tested in the category, all studied categories were *tested* an equal number of times. Specifically, three target items and three lures from each studied category were tested; each novel category was also tested on three occasions. For all categories, 6 items were first randomly selected to serve as the ‘‘critical’’ or to-be-tested target and lure items. Within each category, these 6 items were then further divided into 2 subsets of 3 items each (Subset 1 and Subset 2) which comprised the actual subsets of items that were rotated through the different counterbalancing conditions (studied, nonstudied related, and/or nonstudied novel). For example, if a given participant studied Set A, then the test list would be comprised of the 3 studied target items from each of these categories (A1), 3 within-category lures from these categories that were not, in fact, presented (A2), and 3 novel-category lures from each of the nonstudied categories (B1). In addition, the previously studied 15 unrelated items would be tested (from the A set) as well as 15 unrelated items (lures) that had not been presented (from the B set).

The stimuli were assigned to the study and test lists such that items from the various categories and conditions were distributed throughout the lists, and such that no more than 2 items from any one category ever occurred consecutively. In addition, the test lists were constructed such that no more than 3 previously studied items or new items occurred consecutively.

The study list was comprised of a total of 142 items: 6 items from each of 4 small categories (24 items), 9 items from each of 3 medium categories (27 items), 18 items from each of 4 large categories (72 items), 15 unrelated items, and 4 buffer items (2 presented at the beginning, and 2 presented at the end, of the study list). The test list was comprised of 120 items: 12 small-category targets (3 items from each of 4 categories), 12 small-category related lures (3 items × 4 categories), 6 small-category novel lures (3 items × 2 categories), 9 medium-category targets (3 items × 3 categories), 9 medium-category related lures (3 items × 3 categories), 6 medium-category novel lures (3 items × 2 categories), 12 large-category targets (3 items × 4 categories), 12 large-category related lures (3 items × 4 categories), 12 large-category novel items (3 items × 4 categories), 15 unrelated targets, and 15 unrelated lures.

Procedure. On their first visit to the laboratory, participants were given instructions concerning the picture liking encoding task. They were told that they would see a large number of colored pictures and that their task was to rate how much they liked each picture. It was emphasized that they should determine how much they liked each individual picture, rather than assigning a liking rating for the general class or group of things to which the item belonged. The pictures were then presented by the experimenter, one at a time. After each picture, participants recorded their liking ratings on an appropriately numbered and labeled response form. Liking ratings were given on a 5-point scale, with ''1'' indicating that the participant did not at all like, and ''5'' that they very much liked, the picture. Four different study presentation orders were used: one for each of the four counterbalancing conditions determined by whether Set A1, Set A2, Set B1, or Set B2 was presented at study.

Three days later, on their return visit to the laboratory, participants were given a yes/no recognition test for some of the pictures that they had been shown on their earlier visit. Participants were told that they would again be shown pictures on index cards, one at a time, and that for each picture they were to decide if it was ''old'' (i.e., had been presented previously) or ''new'' (i.e., had not been presented previously). They were asked to indicate their responses by circling either ''old'' or ''new'' on an appropriately labeled response sheet. They were also asked to indicate their confidence in each response, using a 5-point scale where ''1'' indicated that they were just guessing, and ''5'' indicated that they were very sure of their response. As for the study lists, four different test presentation orders were used, corresponding to A1, A2, B1, and B2 as the study conditions.

Results

For each of the experiments reported here, the overall pattern of results is first presented and discussed in terms of hit rates (correct recognition) and false alarms. This is followed by analyses performed on estimates of sensitivity and response bias. We chose to use A' as an estimate of sensitivity and $B_{D''}$ as an estimate of response criteria (Grier, 1971; also see Hodos, 1970; Pollack & Norman, 1964) because values of A' (which provide an estimate of the area under the isosensitivity or ROC curve) have been shown to result in less error than values of d' (which estimate sensitivity in terms of the distance between the means of the distributions associated with two stimulus classes), under conditions where performance shows some bias (Donaldson, 1993). Given the many within-group and between-group conditions in our experiments, it did not seem safe to assume that unbiased performance would uniformly be found. 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). 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 where individuals showed below chance sensitivity (hits $<$ false alarms, or $A' < .50$), the modified formulas for calculating sensitivity and bias for below chance performance (Aaronson & Watts, 1987) were used.

Table 1 presents the mean proportion of hits and false alarms separately by age (old, young), category type (unrelated, within category, or novel category items) and category size (small, medium, large).

First considering correct recognition responses or hits, Table 1 shows that, for the categorized items, older and younger adults showed generally similar rates of correct recognition, with the two age groups achieving average hit rates of 76 and 78%, respectively, for the small-, medium-, and large-category items combined. However, for the unrelated items, older adults showed substantial impairment (68%) compared to their younger counterparts (82%). Next considering false alarms to within-category lures, Table 1 shows that: (a) both age groups more often falsely recognized new within-category items if they had been shown 18 exemplars of the category at study (large-category items) than if they had seen 6 or 9 related items (small- and medium-category items), but (b) the overall level of false alarms was considerably higher in older than in younger adults. For example, whereas in the large-category condition, older adults falsely recognized 70% of the within-category lures—a rate of false recognition that essentially matches the level of *correct* recognition

TABLE 1

Hits and False Alarms, Experiment 1

	Hits				Within category false alarms			Novel category false alarms			
	Unrelated (1)	Small (6)	Medium (9)	Large (18)	Small (6)	Medium (9)	Large (18)	Unrelated (0)	Small (0)	Medium (0)	Large (0)
Old	.68	.73	.72	.83	.43	.40	.70	.11	.21	.05	.08
Young	.82	.75	.78	.81	.21	.25	.35	.03	.03	.02	.05

Note. The values in parentheses indicate the number of related exemplars that were presented at study.

they demonstrated for the small- and medium-category items—younger adults falsely recognized only 35% of the items from large categories. Finally, Table 1 also shows that false alarms to novel-category items and to unrelated items were comparatively infrequent. Young adults showed essentially equivalent rates of novel-category false alarms for all category sizes (average level of 3%), whereas older adults showed higher and more variable baseline rates of false alarms (average of 11%), with especially elevated false alarms to ''small'' category items (21%). Initial analyses performed on the average correct recognition rates for the small, medium, and large categories combined showed no effect of age ($F < 1$), but a significant effect of age for unrelated items with older adults showing significantly depressed recognition relative to their younger counterparts, $F(1,30) = 4.40$, $MS_e = .04$, $p = .04$; older adults also showed significantly elevated rates of false recognition compared to younger adults for all three category sizes: small, $F(1,30) = 13.34$, $MS_e = .03$, $p = .001$; medium, $F(1,30) = 4.99$, $MS_e = .04$, $p = .03$; large, $F(1,30) = 22.82$, $MS_e = .04$, $p < .0001$.

Table 2 presents measures of Sensitivity (A′) and response bias ($B_{D''}$) calculated for the various conditions shown in Table 1. Three measures of sensitivity and bias were computed, using the individual participants' data. The uppermost portion of Table 2 shows estimates of sensitivity and bias comparing hits to novel false alarms (measures of item specific memory, referred to as *A′-Novel* and *$B_{D''}$-Novel*, respectively); the middle portion shows estimates of sensitivity and bias comparing hits to related or within-category false alarms (also measures of item specific memory, re-

TABLE 2

Measures of Sensitivity and Response Bias, Experiment 1

Item specific memory (hits compared to novel false alarms)

Condition	Old A′	Old $B_{D''}$	Young A′	Young $B_{D''}$
Unrelated	.86	.39	.93	.46
Small (6)	.80	.20	.88	.44
Medium (9)	.87	.51	.90	.49
Large (18)	.91	.22	.91	.41

Item specific memory (hits compared to related false alarms)

Condition	Old A′	Old $B_{D''}$	Young A′	Young $B_{D''}$
Small (6)	.72	−.26	.83	.05
Medium (9)	.71	−.20	.82	−.02
Large (18)	.62	−.64	.81	−.30

Gist memory (related false alarms compared to novel false alarms)

Condition	Old A′	Old $B_{D''}$	Young A′	Young $B_{D''}$
Small (6)	.65	.49	.64	.77
Medium (9)	.75	.79	.68	.84
Large (18)	.87	.47	.74	.88

ferred to as *A′-Related* and *$B_{D''}$-Related*, respectively). Finally, the bottom portion of the table presents results when false alarms to related items are treated as a form of gistlike memory, and thus are treated as ’’hits‘‘; here, related false alarms are compared to novel false alarms (cf. Tussing & Greene, in press). These measures are referred to as *A′-Gist and $B_{D''}$-Gist*, respectively.

Each of these indices of sensitivity and response bias was treated as a dependent variable in separate analyses of variance (ANOVAs), treating age (old, young) as a between-subjects factor and category type (unrelated, small, medium, and large; or small, medium, and large) as a within-subjects factor. The outcome of these analyses are reported next, grouped by type of measure.

Item-Specific Memory (Hits Compared to Novel False Alarms)

First, consider the A′ measures shown in the uppermost portion of Table 2, comparing hits to novel false alarms. Whereas older and younger adults showed equal levels of sensitivity for items from large categories, older adults showed comparatively less sensitivity than did younger adults for items from the small and unrelated categories. Consistent with these numerical patterns, a 2 (age) × 4 (category type, including unrelated, small, medium, large) ANOVA on the A′-Novel measure revealed a significant overall effect of age, with younger adults showing higher overall sensitivity (.91) than older adults (.86), $F(1,30) = 5.81$, $MS_e = .01$, $p = .02$. There was also an overall effect of category type, $F(3,90) = 6.59$, $MS_e = .004$, $p = .0004$ and, importantly, a significant interaction of age with category type, $F(3,90) = 2.69$, $MS_e = .004$, $p = .05$. The main effect of category type reflected relatively lower sensitivity for items from small categories than for other items; however, as can be seen from Table 2, this main effect was largely attributable to depressed sensitivity of older adults for the small-category items. A one-way analysis of the A′-novel measure restricted to younger adults showed no effect of category type on sensitivity, $F(3, 45) < 1.7$, whereas a similar analysis considering only older adults showed a significant effect of category type, $F(3,45) = 6.52$, $MS_e = .005$, $p = .0009$. Additional paired *t*-tests comparing the different category types for older adults alone indicated that older adults showed greater sensitivity for the large than for the small [$t(15) = 3.55$, $p = .003$], medium [$t(15) = 2.81$, $p = .01$], or unrelated [$t(15) = 3.43$, $p = .004$] categories; a similar difference was found for the medium compared to the small category [$t(15) = 2.29$, $p = .04$].

We next considered measures of response bias for this same measure (hits compared to novel false alarms). Examination of the means in Table 2 suggests that, for all of the conditions, participants employed comparatively conservative criteria (B_D'' is positive in all cases) but: (a) both older and younger adults showed *relatively* less stringent criteria for large-category items and (b) older adults also showed relatively less stringent criteria for the small-category items. A 2 (age) × 4 (category type: unrelated, small, medium, large) ANOVA performed on these scores showed no overall effect of age, $F < 1$, a trend toward an effect of category type, $F(3,90) = 2.22$, $MS_e = .12$, $p = .09$, and no interaction, $F < 1$. Additional pairwise comparisons treating age as a between-category factor indicated that both age groups employed less stringent criteria for large- than for medium-category items, $F(1,30) = 4.41$, $MS_e = .13$, $p = .04$, and less stringent criteria for small- than for medium-category items, $F(1,30) = 3.67$, $MS_e = .15$, $p = .06$.

Item-Specific Memory (Hits Compared to Related False Alarms)

A two-way analysis examining the effects of age (old/young) and category type (small, medium, large) on the measure of sensitivity comparing hits against related false alarms (A′-Related), revealed strong and consistently greater overall sensitivity among younger (.82) than among older (.68) adults, $F(1, 30) = 31.84$, $MS_e = .01$, $p < .0001$, together with a main effect of category size, $F(1,30) = 3.42$,

MS_e = .01, p = .04, and no interaction, $F <$ 1.7. The effect of category size reflected lower sensitivity, averaging across the two age groups, for items from large categories (.71) than for items from small (.78) or medium (.76) categories but, as can be seen in Table 2, this was primarily attributable to older respondents. Consistent with this, a one-way analysis performed on the A′-Related scores for young adults alone showed no effect of category type, $F < 1$, whereas a one-way ANOVA confined to the older adults showed a significant effect of category type, $F(2,30)$ = 3.92, MS_e = .01, p = .03. Subsequent pairwise comparisons for older adults alone indicated that older adults showed greater sensitivity for small- than for large-category items, $t(15)$ = 2.46, p = .03, and also for medium- than for large-category items, $t(15)$ = 2.13, p = .05.

A 2 (age) × 3 (category type) ANOVA on the corresponding measure of response bias revealed that this age-related decrement in sensitivity was also accompanied by a significant overall difference in response criteria, with older adults using a more lenient criterion overall (M = −.37) than younger adults (M = −.09), $F(1,30)$ = 4.10, MS_e = .45, p = .05. In addition, both older and younger adults employed a relatively more lenient criterion for items drawn from large categories than for small- or medium-category items (overall means of −.47, −.10, and −.11, respectively), $F(2, 60)$ = 11.62, MS_e = .12, $p < .0001$, but the degree to which this was true did not interact with age, $F < 1$. Pairwise comparisons showed that participants used significantly less stringent criteria for large than for either small, $F(1,30)$ = 13.39, MS_e = .16, p = .001, or medium, $F(1,30)$ = 24.99, MS_e = .08, $p < .0001$, categories.

Gist Memory (Related False Alarms Compared to Novel False Alarms)

Levels of A′ when related false alarms are treated as ‘‘hits’’ do not necessarily indicate differences in the level of gist memory that was available to participants, but do reflect differences in the degree to which such memory was actually employed in participants’ recognition judgments, despite any countervailing forces of item-specific memory. A 2 (age) × 3 (category type: small, medium, large) ANOVA performed on this measure revealed a trend toward an overall effect of age, $F(1,30)$ = 3.36, MS_e = .03, p = .08, a main effect of category type, $F(2,60)$ = 15.34, MS_e = .01, $p < .0001$, and a trend toward an interaction of age with category type, $F(2,60)$ = 2.38, MS_e = .01, p = .10. Overall ‘‘gistlike’’ sensitivity was higher for large- and medium-category items (.80 and .72, respectively) than for small-category items (.64), but this difference was more pronounced for older adults (.87, .75, .65, respectively) than for younger adults (.74, .68, .64, respectively). Pairwise comparisons showed greater gistlike sensitivity for items from large than small, $F(1,30)$ = 28.71, MS_e = .01, $p < .0001$, large than medium, $F(1,30)$ = 13.54, MS_e = .009, p = .0009, and medium than small, $F(1,30)$ = 5.11, MS_e = .02, p = .03, categories; there was also a significant age × category type interaction for the large versus small comparison, $F(1,30)$ = 4.46, MS_e = .01, p = .04. A one-way analysis considering the large-category items only revealed that older adults showed significantly greater gist sensitivity for large-category items than did younger adults, $F(1,30)$ = 12.88, MS_e = .01, p = .001.

A 2 (age) × 3 (category type) ANOVA performed on the corresponding index of response criteria ($B_{D''}$) for this measure showed a significant main effect of age, $F(1,30)$ = 8.08, MS_e = .18, p = .008, a main effect of category type, $F(2,60)$ = 4.06, MS_e = .07, p = .02, together with an age × category type interaction, $F(2,60)$ = 3.82, MS_e = .07, p = .03. Although both older and younger adults employed relatively conservative criteria, younger adults were significantly more conservative (.83) than older adults (.58). Also, older individuals were most conservative for medium-category items (.79), with somewhat more lenient responding for small- (.49) and large- (.47) category items, whereas younger adults were less affected by category size (.77, .84, and .88 for small, medium, and large, respectively). Pairwise comparisons revealed consistent effects of age (smallest $F = 4.07$), and a significant age × category interaction for the medium vs large comparison, $F(1,30) = 10.02$, $MS_e = .05$, $p = .004$.

High-Confidence Responses Only

Participants were also asked to indicate their level of confidence in the recognition responses they gave using a 5-point scale, where ''5'' indicated they were ''very sure'' and ''1'' indicated they were ''just guessing.''[1] Restricting consideration to only those responses that were accompanied by high confidence (responses of ''4'' or ''5''), the hit rates for older adults for unrelated items and for small-, medium-, and large-category items were 45, 52, 43, and 68%, respectively; high confidence hits for younger adults were 66, 55, 58, and 59%, respectively. Thus the age-related deficit in correct recognition of unrelated items was even more pronounced on this measure (deficit of 21%) than in overall recognition (deficit of 14%).

In contrast to their manifested willingness to give high confidence ratings to their *correct* recognition responses, younger adults rarely professed high confidence in the items they falsely recognized, whereas older adults relatively often did so—especially for large categories. High-confidence false recognition rates for older adults for unrelated and small-, medium-, and large-category items were 5, 29, 19, and 50%, respectively; the corresponding values for younger adults were <1, 10, 10, and 16%. Very few false alarms to novel-category items were accompanied by high confidence (for older adults, 13, 2, and 3% for the small, medium, and large nonstudied categories, respectively; for young adults, <1% in all cases).

[1] Given that we collected confidence ratings from participants, ROC curves might have been calculated directly rather than using estimates of A′. However, the total number of observations per participant in each condition was small (typically 15 in Experiments 2 and 3; fewer than this for some conditions of Experiment 1) and, to obtain ROC curves, these observations would need to be further subdivided on the basis of the participant's old/new responses as well as their confidence ratings. For these reasons, we used the nonparametric estimates of A′ and $B_{D''}$ as our primary dependent measures for statistical analysis; however, for descriptive purposes, we also report the numerical proportions of correct recognition responses and false alarms for comparatively higher levels of confidence (combining across the two highest ratings of confidence).

Discussion

This experiment produced three particularly notable findings. First, relative to the level of false recognition shown for novel items, both younger and older adults showed substantial false recognition of detailed colored pictures, thus extending the range of robust false recognition beyond previous studies of words and line drawings. For younger adults, false alarms to within-category lures were, on average, nine times more frequent ($M = .27$) than false alarms to novel-category lures ($M = .03$); for older adults, who also showed a higher level of baseline false alarms, false alarms to within category lures ($M = .51$) were five times more frequent than false alarms to novel-category items ($M = .11$). For older adults, false recognition was especially pronounced for within-category lures when a large number of similar items had been encountered, such that older adults falsely recognized within-category lures from large categories almost as often (70%) as they correctly recognized items from small and medium categories (73 and 72%) or unrelated items (68%). Second, signal detection analyses revealed that older and younger adults did not show a reliable overall difference in response criteria when hits were compared to novel false alarms, but older adults were consistently more lenient than younger adults when hits were compared to related false alarms. However, the latter criterion difference was *also* accompanied by a reliable age difference in sensitivity, such that older adults were consistently less able to discriminate studied items from related lures than were younger adults. Third, signal detection analyses also revealed that sensitivity of older adults for hits compared to novel false alarms was greatest for medium and large categories and less for small and unrelated items. Together, these findings suggest that older adults were relying, to a greater extent than were younger adults, on a generic or gistlike representation of the studied items. This age-related differential reliance on generic representations was not associated with any particular cost when hits were compared to novel lures, as seen in the high level of sensitivity of older adults for medium- and large-category items and the nearly equivalent sensitivity of old and young (averages of .89 vs .91, respectively). However, it was associated with a cost when hits were compared to related lures, where older adults showed more strongly impaired sensitivity relative to the young (averages of .67 versus .82 for medium and large categories combined for older versus younger respondents, respectively).

Thus, older adults showed a significantly increased susceptibility to false recognition of within-category lures in a paradigm that was intended to eliminate ''source confusions'' (in the strict sense noted previously) as a potential reason for the increment in age-related false recognition. Although within-category lures shared many perceptual and conceptual features with the studied items, it is very improbable that these lures were themselves generated or imagined by the participants during their initial encounter with the study stimuli—as, for example, inferences might be generated during the reading of a passage of text, or the critical lure word might be produced or ''spring to mind'' in the converging associates paradigm used by Deese (1959), Roediger and McDermott (1995), and others. Rather, erroneous recognition of within-category lures presumably occurred not because participants confused the true origins of their experiences (the specific picture for which they were asked to make a recognition judgment had never been presented and so a source confusion, in a strict sense, was almost impossible), but because they were unable to (or simply did not) differentiate between exemplars that had been presented and other—quite similar—exemplars that had not been presented but that they now mistakenly claimed they had seen.

Nonetheless, consideration of the experimental design of the present experiment raises an important caveat that—in the absence of any further data—would seriously undermine the generality of the foregoing conclusions. This caveat concerns the manner in which stimuli were assigned to the ''study category'' conditions. Although within a given category size, the particular categories and the items that comprised studied and nonstudied categories were counterbalanced across participants (this was true for 16 of the 19 categories used), the categories themselves were not counterbalanced across the *category size* manipulation. If, for example, cats or dinosaurs were studied, exemplars from these categories were always studied as part of a large category; similarly, medium-category items were always medium-category items and small-category items were always small-category items. Stated differently, category size was confounded with stimulus type. This confound derived from limitations on the numbers of categorized pictures that were available when we carried out the experiment, but could be avoided if additional categories of pictures, all with a large number of exemplars, could be obtained; any one category could then serve in all category size conditions by the simple expedient of presenting either a greater or smaller number of the exemplars from each category. This was the approach adopted in the next experiment, where category sizes of 9 vs 18 exemplars each were examined; and in Experiment 3, where category sizes of 1, 9, and 18 were considered.

EXPERIMENT 2

Method

Participants. Older and younger individuals were recruited and screened according to the same neuropsychological and other criteria as in Experiment 1. Sixteen older adults (*M* age = 68.5 yrs, range = 64–74) and 16 younger adults (*M* age = 19.8, range = 18–24) took part. Older adults had on average 15.0 years of formal education (range = 12–22 yrs); younger adults had on average 13.9 years (range = 12–18). All participants were native speakers of English and had normal or corrected-to-normal vision. Participants were paid for their involvement in the study.

Procedure. The general experimental procedure was similar to Experiment 1: participants initially were shown colored pictures, rated their liking for the pictures on a 5- point scale, and then, 3 days later, were given a yes/no recognition test. However, significant changes in the procedure occurred in relation to (a) the category composition of the stimuli and (b) the mode of experimental presentation.

The particular stimuli used were changed: additional colored pictures were obtained from various illustrated books for children and adults, such that a set of 20 categories, each comprised of a total of 21 exemplars, was created. These categories were divided into 4 subsets of 5 categories each, and the subsets were then rotated across the experimental manipulation of category size, such that sometimes a given subset of 5 categories (e.g., Set P, comprised of cars, cats, children, clocks, and flowers, or Set Q, comprised of birds, shelves, teapots, teddy bears, and whales) served as a large category (18 exemplars of each category presented at study), sometimes as a medium category (9 exemplars of each category presented at study), and sometimes was not presented, to serve as baseline measure of ‘‘novel category’’ false alarms (here, two of the category subsets, or a total of 10 categories, were used). In addition, 30 unrelated items, (15 presented both at study and at test and 15 presented as unrelated lures at test) were used as in the previous experiment.

The mode of presentation was also changed, such that both study and test stimuli were presented via computer.

All stimuli were first mounted on plain white index cards and then scanned and converted to digital format using VistaScan and a UMAX Vista-S6E scanner. For study and test presentation, the stimuli were displayed in the center of a color computer monitor, using a 256-color look-up table, and the PsyScope experimental presentation program. As in Experiment 1, during the study phase, pictures were presented for 2 s each, and participants performed the liking rating encoding task immediately after each item. However, now participants entered their liking ratings via the keyboard rather than on written response forms. After each item, they received a prompt asking them to make their liking rating; participants entered their rating and then pressed the ‘‘tab’’ key when they were ready to proceed to the next item. During the test phase, the stimuli were presented at the center of the monitor, and beneath the stimulus two separate prompts appeared requesting participants to enter an Old/New response (‘‘Old or New?,’’ displayed at the bottom left of the screen), and then their degree of confidence (‘‘Confidence? 1 = just guessing, 5 = very sure,’’ displayed at the bottom right of the screen).

The study list was comprised of a total of 156 items, including 90 large-category items (18 items from each of 5 categories), 45 medium-category items (9 items from each of 5 categories), 15 unrelated items, and 6 buffer items (3 presented at the beginning and 3 presented at the end of the study list). The test list included 3 old items from each of the studied categories (15 large-category targets, 15 medium-category targets), 3 new items from each studied category (15 large-category related lures, 15 medium-category related lures), and 3 new items from each of 10 nonstudied categories (i.e., 30 novel category lures), as well as 15 unrelated targets and 15 unrelated new items.

Results

Table 3 presents hits and false alarms. Examination of hits reveals a pattern of performance similar to that observed in Experiment 1: The two age groups showed comparatively similar levels of correct recognition for the items from the medium and large object categories (averages of 77 vs 82% for old and young, respectively), but older adults showed a considerably more pronounced decrement in correct recognition for the unrelated items (averages of 66 vs 86%, respectively). The pattern of within-category false alarms also appears fairly similar to that observed in Experiment 1, with older adults showing a higher

TABLE 3

Hits and False Alarms, Experiment 2

	Hits			Within category false alarms		Novel category false alarms	
	Unrelated (1)	Medium (9)	Large (18)	Medium (9)	Large (18)	Unrelated (0)	Categorized (0)
Old	.66	.76	.78	.50	.58	.13	.10
Young	.86	.81	.83	.23	.26	.04	.05

Note. The values in parentheses indicate the number of related exemplars that were presented at study.

level of within-category false alarms than younger adults (averages of 54 vs 25%, respectively) and also more strongly influenced by the number of related exemplars presented at study than their younger counterparts (related false alarms of 50 and 58% for medium and large categories, respectively, compared to 23 and 26% for younger adults). Finally, novel-category false alarms also show a broadly similar pattern to Experiment 1, with older adults showing baseline false alarm rates of between 10 and 13%, compared to a rate of 4 and 5% for younger adults. Initial analyses showed no age difference in correct recognition of items from the medium and large categories combined ($F < 1.3$), but a significant age-related decrement in correct recognition of unrelated items, $F(1,30) = 18.64$, $MS_e = .02$, $p = .0002$. Analyses of false alarms to within-category lures showed elevated rates of false recognition for older compared to younger adults for both medium- and large-category items, $F(1,30) = 19.64$, $MS_e = .03$, $p = .0001$; $F(1,30) = 23.07$, $MS_e = .03$, $p < .0001$, respectively.

Table 4 presents the measures of Sensitivity (A′) and response bias ($B_{D''}$) for the various conditions. The same three measures of sensitivity and bias that were computed in Experiment 1 are again considered, with each measure treated as a dependent variable in separate two-way analyses of variance (ANOVA), entering age (old, young) as a between-subjects factor and category type (unrelated, medium, and large, or medium and large) as a within-subjects factor.

Item-Specific Memory (Hits Compared to Novel False Alarms)

A 2 (age) × 3 (category type: unrelated, medium, large) ANOVA performed on the measure of sensitivity comparing hits to novel

TABLE 4

Measures of Sensitivity and Response Bias, Experiment 2

Item specific memory (hits compared to novel false alarms)

	Old		Young	
Condition	A′	$B_{D''}$	A′	$B_{D''}$
Unrelated	.84	.46	.94	.31
Medium (9)	.89	.40	.93	.48
Large (18)	.90	.34	.93	.44

Item specific memory (hits compared to related false alarms)

	Old		Young	
Condition	A′	$B_{D''}$	A′	$B_{D''}$
Medium (9)	.71	−.41	.86	−.11
Large (18)	.67	−.50	.85	−.23

Gist memory (related false alarms compared to novel false alarms)

	Old		Young	
Condition	A′	$B_{D''}$	A′	$B_{D''}$
Medium (9)	.80	.71	.72	.94
Large (18)	.83	.62	.73	.88

category false alarms revealed a significant main effect of age, $F(1,30) = 23.23$, $MS_e = .003$, $p < .0001$, a main effect of category type, $F(2,60) = 3.77$, $MS_e = .002$, $p = .03$, and a significant age × category type interaction, $F(2,60) = 5.83$, $MS_e = .002$, $p = .005$. Younger adults showed greater overall sensitivity (.93) than did older adults (.88), and older but not younger participants showed decreased sensitivity for unrelated items (.84 vs .94, respectively). Separate pairwise comparisons of the category types showed a consistent effect of age (smallest $F = 6.94$), and a significant age × category type interaction in both the unrelated vs medium-category comparison, $F(1,30) = 6.94$, $MS_e = .002$, $p = .01$, and the unrelated vs large-category comparison, $F(1,30) = 6.79$, $MS_e = .002$, $p = .01$.

A 2 (age) × 3 (category type: unrelated, medium, large) ANOVA performed on the corresponding measure of response criteria indicated that there were no differences in response criteria as a function of age or category type (Fs < 1) and no interaction ($F < 1.8$).

Item-Specific Memory (Hits Compared to Related False Alarms)

A 2 × 2 ANOVA examining the effects of age and category type (medium, large) on sensitivity when hits were compared against within-category lures (A′-Related), revealed that younger adults showed substantially greater sensitivity (.86) than did older adults (.69), $F(1,30) = 43.70$, $MS_e = .01$, $p < .0001$. There was no effect of category type on this measure ($F < 1.3$) and no interaction of age with category type ($F < 1$).

Although older adults showed a numerical tendency toward more lenient responding ($B_{D''} = -.46$) than did younger adults ($B_{D''} = -.17$), a 2 (age) × 2 (category type: medium, large) ANOVA performed on the response-criterion measure revealed that this difference was not reliable, $F(1,30) = 2.46$, $MS_e = .54$, $p = .13$. Both older and younger adults were slightly, but not significantly, more lenient in responding to large- than to medium-category items, $F(1,30) = 2.41$, $MS_e = .08$, $p = .13$.

Gist Memory (Related False Alarms Compared to Novel False Alarms)

A two-way ANOVA, treating age (old, young) as a between- and category type (medium or large) as a within-subjects factor was also performed on the A′ measures derived when treating related false alarms as hits compared to novel-category false alarms (i.e., A′-Gist). This analysis revealed a significant main effect of age, $F(1,30) = 11.10$, $MS_e = .01$, $p = .002$, and a slight trend toward an effect of category type, $F(1,30) = 2.40$, $MS_e = .002$, $p = .13$. On this measure, which could be viewed as an index of willingness to act on gistlike memory, older adults showed higher sensitivity (.82) than did younger adults (.73), who may have used more item-specific memory to oppose such a tendency. Across the two age groups, sensitivity to gist was slightly greater for exemplars from 18-item categories (.78) than for items from 9-item categories (.76).

On the corresponding measure of response bias, both older and younger adults showed conservative tendencies, but younger adults did so to a significantly greater extent (.91) than older adults (.67), $F(1,30) = 8.18$, $MS_e = .12$, $p = .008$. Both older and younger participants showed greater leniency for the large-category than for the medium-category items, $F(1,30) = 4.88$, $MS_e = .02$, $p = .03$; $F < 1$ for the category × age interaction.

High Confidence Responses Only

The proportions of hits and false alarms shown by older and younger adults were also examined when considering only those positive recognition responses that were accompanied by high confidence (''4'' or ''5'' on a 5-point scale, where ''5'' indicated that participants were ''very sure'' and ''1'' indicated they were ''just guessing''). As in Experiment 1, older adults also showed depressed recognition of unrelated items compared to their level of recognition of the categorized items on this measure. Older adults' high confidence hits for the unrelated, medium- and large-category items were 45, 64, and 62%, respectively; the corresponding values for younger adults were 68, 56, and 58%. High confidence false recognition of within-category lures also showed a pattern similar to that found in Experiment 1, with older adults more often providing high confidence false recognition responses, particularly for the related lures from large categories, and younger adults relatively seldom doing so (older adults: rates of 5, 4, 37, and 42% for unrelated items, novel items, and medium- and large-category items, respectively; younger adults: rates of 2, 1, 4, and 8%, respectively).

Discussion

This experiment has replicated many of the central findings from Experiment 1. Both older and younger adults showed substantial rates of false recognition of pictorial lure items that were related to items that they had studied. For large-category items (for which 18 similar exemplars were presented at study), older adults were nearly six times as likely to falsely claim to recognize within-category lures than to false alarm to baseline ''novel'' items (.58 vs .10); younger adults were approximately five times as likely to do so (.26 vs .05). Older adults also again demonstrated a notable impoverishment of *correct* recognition for items that were unrelated to categorized items, together with distinctly less marked deficits in recognition of items from medium and large categories, as reflected in significant interactions of age with category type for the A′-Novel measure of sensitivity, both when sensitivity for medium- and sensitivity for large-category items was contrasted with that for unrelated items. These differences in sensitivity were unaccompanied by differences in response criteria. However, also as in Experiment 1, age-related differences in sensitivity were more pronounced when hits were compared to within-category lures; here the average sensitivity (A′) of older adults (.69) was considerably less than that of younger adults (.86). Although older adults in Experiment 1 also showed more lenient response criteria than did younger adults, in this experiment the latter differences in sensitivity were unaccompanied by significant age differences in response criteria.

EXPERIMENT 3

Although Experiment 2 replicated the major features of Experiment 1, there was one notable difference: Whereas, in Experiment 1, younger adults showed some increases in false alarms for categories of 6, 9, and 18 pictures (false alarms of 21, 25, and 35%, respectively), they showed no such increases in Experiment 2 with categories of 9 and 18 pictures (false alarms of 23 and 26%, respectively). This latter finding contrasts with data reported by Robinson and Roediger (1997), using the Deese/Roediger–McDermott converging associates paradigm, who found more false alarms to related lure words with the presentation of increasing numbers of associates during study. It is possible that we failed to observe increased false recognition as a function of the number of related pictures shown at study with younger adults in Experiment 2 because we sampled from a relatively restricted range (i.e., categories were represented by 9 vs 18 pictures). To address this issue, in Experiment 3 we included a condition in which only a single picture from a category was presented in the study list and examined false alarms to a single related lure picture on the recognition test.

Including categorized items for which only one exemplar was presented at study also allowed us to examine whether the low hit rate by older adults to unrelated pictures in Experiments 1 and 2 reflects a systematic difference between the specific items that served as unrelated pictures and the specific items that served as categorized pictures. For example, the categorized items may have been named more readily or may have had greater extra-experimental familiarity than the unrelated items. In Experiment 3, categorized pictures were rotated through the three different category sizes (1, 9, and 18). By comparing the recognition hit rates for single categorized pictures to those found for unrelated pictures (the unrelated pictures used in Experiments 1 and 2 were again used in Experiment 3), we can determine whether or not the relevant age differences observed in the first two experiments are specific to the particular pictures that serve as unrelated items.

To accomplish these objectives, we had to reduce the number of items used to estimate novel category false alarms from 10 categories (as in Experiment 2) to 5 categories, which could increase idiosyncratic differences in responding to new items (particularly for older adults who, on the basis of Experiment 1 and pilot studies, tended to respond somewhat more variably than younger adults to novel category items). Nevertheless, the advantages of including the single categorized picture condition more than offset the potential costs.

Method

Participants. Older and younger participants were recruited and screened according to the same neuropsychological and other criteria employed in the previous experiments. Sixteen older (*M* age = 70.4 yrs, range = 64–75) and 16 younger (*M* age = 19.7, range = 18–25) adults took part. Older adults had, on average, 14.4 years of formal education (range = 12–20); younger adults had on average 13.9 years (range = 12–19).

Procedure. With two exceptions, the procedure was identical to that for Experiment 2. The first alteration concerned the addition of single item categories. In addition to medium (9-exemplar) and large (18-exemplar) study categories, a single or 1-item study category was included, and the number of categories used to estimate novel false alarms was reduced from 10 categories to 5 categories. Specifically, the 20 categories were divided into 4 sets of 5 categories each (Sets P, Q, R, and S). Across subjects, these sets were rotated through the different category-size conditions (nonstudied or novel, single, medium, and large), such that each set of categories equally often served in each condition, and such that the critical (tested) subsets of items within each category (Subsets 1 and 2) equally often occurred as studied items (targets) and as nonstudied items (lures). (For the single-item condition, a reduced subset of 2 items from the 6 ‘‘critical’’ items in each category was randomly selected for inclusion as targets and lures; this reduced subset was then rotated through the studied and nonstudied conditions in the same manner as other subset items.) In all, the counterbalancing procedure necessitated the construction of 16 different study lists and 16 different test lists. One older adult and one young adult were assigned to each of these study-test combinations.

The second procedural alteration concerned the number of items included in the study list. The study list in the present experiment was increased to 215 items, compared to 156 items in Experiment 2. This difference resulted from the addition of the 5 ‘‘single’’ studied items to the study list and from the inclusion of 54 additional filler items. These nontested filler items were drawn from 4 categories that were unrelated to the other categories and were comprised of two 9-item (‘‘medium’’) categories and two 18-item (‘‘large’’) categories. These filler items were randomly interspersed throughout the study list in the same manner as the actual study items and were included to increase the study list length, with the aim of increasing the baseline rate of false alarms to novel category items in younger adults. However, as will be seen, younger adults showed an identical baseline rate of false alarms (5%) to that observed in Experiment 2.

In summary, then, the study list was comprised of 215 items, including 5 single items (1 item from each of 5 object categories), 45 medium-category items (9 items from each of 5 categories), 90 large-category items (18 exemplars from each of 5 categories), 15 unrelated items, 54 filler items, and 6 buffer items. The test list was comprised of 115 items: 5 single targets, 5 single related lures, 15 medium-category targets, 15 medium-category related lures, 15 large-category targets, 15 large-category related lures, 15 unrelated targets, 15 unrelated lures, and 15 novel-category lures.

Results

Table 5 presents hits and false alarms; Table 6 presents measures of sensitivity and response criteria.

TABLE 5

Hits and False Alarms, Experiment 3

	Hits				Within category false alarms			Novel category false alarms	
	Unrelated (1)	Single (1)	Medium (9)	Large (18)	Single (1)	Medium (9)	Large (18)	Unrelated (0)	Categorized (0)
Old	.68	.68	.79	.83	.29	.51	.63	.13	.22
Young	.88	.90	.81	.81	.11	.28	.25	.06	.05

Note. The values in parentheses indicate the number of related exemplars that were presented at study.

First, consider correct recognition. From Table 5, we see that hits for the medium- and large-category items showed no differences as a function of age, with older and younger adults achieving an identical average hit rate of 81% for these categories. However, compared to younger adults, older adults again showed considerably reduced correct recognition for unrelated items (68 vs 88%). Critically, this difference was virtually identical to that observed for the single items (68 vs 90%), suggesting that unrelated items and single items were functionally similar to one another and arguing against the notion that the depressed recognition of unrelated items observed among older adults in the two previous experiments was attributable to characteristics peculiar to the unrelated items, rather than to the fact that they comprised single exemplars of a particular category. Next considering false alarms, Table 5 shows that the pattern of false alarms for the medium- and large-related category lures was also very consistent with that observed in the two previous experiments, with older adults showing higher levels of false recognition than younger adults (averages of 57 vs 27%), and older adults demonstrating higher rates of false recognition for the large- than for the medium-category lure items (63 vs 51%, respectively), whereas this was not true for younger adults (25 vs 28%, respectively). Importantly, however, both age groups showed an effect of category size on related-item false alarm rates when the category size comprised a *single* item. Compared to the baseline level of 22% for novel items, older adults falsely recognized 29% of items for which a single related exemplar had been shown; the corresponding values for younger adults were 5 and 11%. Thus, whereas—as

TABLE 6

Measures of Sensitivity and Response Bias, Experiment 3

Item specific memory (hits compared to novel false alarms)

	Old		Young	
Condition	A′	$B_{D''}$	A′	$B_{D''}$
Unrelated	.84	.44	.93	.19
Single (1)	.79	.24	.93	.32
Medium (9)	.85	−.002	.92	.46
Large (18)	.87	−.14	.92	.44

Item specific memory (hits compared to related false alarms)

	Old		Young	
Condition	A′	$B_{D''}$	A′	$B_{D''}$
Single (1)	.73	.04	.89	−.03
Medium (9)	.71	−.48	.83	−.16
Large (18)	.67	−.59	.85	−.12

Gist memory (related false alarms compared to novel false alarms)

	Old		Young	
Condition	A′	$B_{D''}$	A′	$B_{D''}$
Single (1)	.59	.50	.63	.65
Medium (9)	.72	.45	.73	.90
Large (18)	.77	.24	.71	.88

also observed in Experiment 2—older adults but not younger adults showed an increased tendency to falsely recognize related lures as a function of the number of exemplars presented at study when only the medium (9-exemplar) and large (18-exemplar) categories were considered, both age groups showed a clear susceptibility to elevated false alarms as a function of category size when contrasting false-recognition of single-item categories with the two larger categories (false alarms of 11 vs an average of 27% for younger adults, and 29 vs 57% for older adults). Initial analyses showed no age differences in correct recognition of items from the medium and large categories combined ($F < 1$), but a significant age-related impairment in correct recognition for the single items, $F(1,30) = 10.21$, $MS_e = .04$, $p = .003$, and the unrelated items, $F(1,30) = 13.61$, $MS_e = .02$, $p = .0009$. Older adults showed higher rates of false recognition than did younger adults for single-item category lures, $F(1,30) = 6.59$, $MS_e = .04$, $p = .02$, and for both medium- and large-category item lures, $F(1,30) = 15.77$, $MS_e = .03$, $p = .0004$, and $F(1,30) = 34.26$, $MS_e = .03$, $p < .0001$, respectively. In addition, both older and younger adults showed significantly greater false alarms to single-category lures than to novel-category lures, $F(1,30) = 5.44$, $MS_e = .01$, $p = .03$; $F < 1$ for the age × category type interaction.

Item-Specific Memory (Hits Compared to Novel False Alarms)

A two-way ANOVA performed on the measure of sensitivity comparing hits to novel false alarms (A′-Novel), treating age as a between-subjects factor and category type (unrelated, single, medium, large) as a within-subjects factor revealed a significant main effect of age, $F(1,30) = 30.48$, $MS_e = .008$, $p < .0001$, and a significant age × category type interaction, $F(3,90) = 3.92$, $MS_e = .004$, $p = .01$. Overall, younger adults showed greater sensitivity (.93) than did older adults (.84), but this difference in sensitivity was particularly apparent for the single items and, to a lesser degree, the unrelated items, such that whereas older adults showed somewhat diminished sensitivity for these items, younger adults showed slightly greater sensitivity for these items (see Table 6). Pairwise comparisons revealed consistent effects of age (smallest $F = 11.33$) and a significant interaction of age with category type for the comparison of medium- with single-category items, $F(1,30) = 7.75$, $MS_e = .003$, $p = .009$, and the comparison of large with single-category items, $F(1,30) = 9.39$, $MS_e = .004$, $p = .005$.

A 2 (age) × 4 (category type: unrelated, single, medium, large) analysis performed on the corresponding measure of response criteria revealed a nonsignificant trend toward an overall effect of age, $F(1,30) = 2.36$, $MS_e = .65$, $p = .13$, and a significant age × category type interaction, $F(3,90) = 8.80$, $MS_e = .13$, $p < .0001$. Both older and younger adults employed comparatively conservative criteria for unrelated and single items, but younger adults also tended to be more conservative than older adults for the medium- and large-category items. Pairwise comparisons revealed a main effect of age when medium and large categories were compared, $F(1,30) = 9.82$, $MS_e = .44$, $p = .004$, and when large and single categories were compared, $F(1,30) = 4.30$, $MS_e = .41$, $p = .05$. These comparisons also revealed several interactions of age with category type reflecting the following patterns: (1) in the comparison of responding to single- vs medium-category items, younger adults were relatively more lenient for single- than for medium-category items, whereas the reverse was true for older adults, $F(1,30) = 9.32$, $MS_e = .06$, $p = .005$; (2) a similar pattern was observed for the comparison of single- vs large-category items: whereas younger adults were relatively more lenient for single- than for large-category items, the reverse was true for older adults, $F(1,30) = 9.84$, $MS_e = .10$, $p = .004$; (3) in the comparison of unrelated to large-category items, older adults showed *relatively* lenient criteria for the large-category items, $F(1,30) = 14.94$, $MS_e = .18$, $p = .0006$; with (4) a similar pattern observed in comparison of unrelated to medium-category items (i.e., older adults responding relatively more leniently for medium-category items), $F(1,30) = 11.77$, $MS_e = .17$, $p = .002$.

Item-Specific Memory (Hits Compared to Related False Alarms)

A two-way ANOVA performed on the A′-Related measure, comparing hits against related false alarms, and including age (old, young) as a between-subjects factor and category type (single, medium, large) as a within-subjects factor, revealed a main effect of age, $F(1,30) = 35.24$, $MS_e = .02$, $p < .0001$, and a trend toward a main effect of category type, $F(2, 60) = 2.23$, $MS_e = .01$, $p = .12$. Younger adults showed greater sensitivity overall (.86) than did older adults (.70), and both older and younger adults demonstrated greater sensitivity for single items (average of .81) than for medium- or large-category items (average of .77 and .76, respectively). Pairwise comparisons showed a consistent effect of age (smallest $F = 21.70$), and trends toward an effect of category type both when performance for single-category items was compared to that for medium-category items, $F(1,30) = 3.46$, $MS_e = .007$, $p = .07$, and when performance for single-category items was compared to that for large-category items, $F(1,30) = 3.30$, $MS_e = .01$, $p = .08$.

A 2 (age) × 3 (category type: single, medium, large) ANOVA performed on the corresponding response criterion measure showed a trend toward an overall effect of age, $F(1,30) = 3.46$, $MS_e = .39$, $p = .07$, a main effect of category type, $F(2,60) = 8.18$, $MS_e = .16$, $p = .0007$, and an age × category type interaction, $F(2, 60) = 3.86$, $MS_e = .16$, $p = .03$. Both older and younger adults employed relatively more conservative criteria for single items (B''_D average near zero) than for medium- or large- category items (B''_D average of −.32 and −.36, respectively). Pairwise comparisons showed a main effect of age when responding to medium- vs large-category items was compared, $F(1,30) = 7.53$, $MS_e = .33$, $p = .01$, but not for the single vs medium, or single vs large, comparisons, Fs < 2.3. These comparisons also showed an effect of category type in the single- vs medium-, and in the single- vs large-category comparisons, reflecting overall more lenient responding to items from the medium and large categories, $F(1,30) = 11.44$, $MS_e = .15$, $p = .002$, and $F(1,30) = 9.98$, $MS_e = .21$, $p = .004$, respectively, but these effects were carried more by older than by younger adults, as reflected in age × category type interactions, $F(1,30) = 3.88$, $MS_e = .15$, $p = .06$, and $F(1,30) = 5.39$, $MS_e = .21$, $p = .03$.

Gist Memory (Related False Alarms Compared to Novel False Alarms)

We again considered the effects of age and category type when A′ was calculated such that false recognition responses to within-category lures were treated as ‘‘hits’’ (indicating gistlike memory) and compared to false alarms to novel category items. A 2 (age) × 3 (category type: single, medium, large) ANOVA performed on this measure revealed a significant effect of category type, $F(2,60) = 12.81$, $MS_e = .01$, $p < .0001$, reflecting greater overall sensitivity to gist for medium- (.73) and large- (.74) category items than for single (.61) items. Whereas older adults showed slightly greater sensitivity to gist (and willingness to respond to gist) for large- (.77) than for medium- (.72) category items, the reverse tended to be true for younger adults, with younger adults more often responding to gist for the medium- (.73) than for the large- (.71) category items. Pairwise comparisons showed main effects of category type, reflecting greater overall sensitivity to gist for both the medium vs single, $F(1,30) = 17.52$, $MS_e = .01$, $p = .0002$, and large vs single, $F(1,30) = 16.38$, $MS_e = .02$, $p = .0003$, comparisons.

A 2 (age) × 3 (category type: single, medium, large) ANOVA on the response criterion measure revealed a main effect of age, $F(1,30) = 14.98$, $MS_e = .27$, $p = .0005$, and an age × category type interaction, $F(2,60) = 5.23$, $MS_e = .09$, $p = .008$. Younger adults exercised more conservative criteria overall (.81) than did older adults (.40), and older adults showed their most lenient responding for items from large categories (.24 vs .50 and .45 for single and medium, respectively) whereas younger adults were most lenient for single items (.65 vs .90 and .88 for medium and large, respectively). Pairwise comparisons revealed consistent main effects of age (smallest $F = 6.76$) and, for the single vs medium and single vs large comparisons, interactions of age with category type, reflecting the relatively more lenient criteria of older adults for medium- and large-category items compared to their own and younger adults' criteria for single-category items, $F(1,30) = 3.78$, $MS_e = .09$, $p = .06$, and $F(1,30) = 7.76$, $MS_e = .12$, $p = .009$, respectively.

High-Confidence Responses Only

The proportions of hits and false alarms shown by older and younger adults considering only those recognition responses that were accompanied by high confidence (''4'' or ''5'' on a 5-point scale, where ''5'' indicated participants were ''very sure'' and ''1'' indicated they were ''just guessing'') showed a similar pattern to that obtained in Experiments 1 and 2. Whereas older adults showed fewer high confidence hits for unrelated items (47%) and also for single items (46%) than for medium- (60%) or large- (63%) category items, younger adults showed, if anything, the reverse pattern (high confidence hit rates of 68, 71, 62, and 63% for unrelated, single-, medium-, and large-category items, respectively). For false-recognition responses, again, older adults showed a considerable proportion of high confidence false recognition, whereas younger adults showed comparatively few such responses (for unrelated, novel-, single-, medium-, and large-category items respectively, older adults: 8, 11, 18, 35, and 46%; younger adults: 4, 2, 3, 10, and 9%, respectively).

Discussion

Many of the key findings from the two earlier experiments were again observed. The pattern of increased false recognition of related lures by older than younger adults was again clearly seen. Also, as in the previous experiments, on measures of sensitivity comparing hits to false alarms for novel categories, younger adults showed strong sensitivity regardless of the number of targets shown at study whereas older adults showed a decrement for items represented by only one item. Critically, the inclusion of items from categories where only one item was presented at study yielded correct recognition rates highly similar to those found with unrelated items, both in the present experiment, and in Experiments 1 and 2, pointing to the generalizability of the finding that older adults showed an especially marked deficit in correct recognition for noncategorized (one-of-a-kind) items. The results of the present experiment also show that even though false recognition responses of younger adults showed little effect of the number of similar items presented at study in the comparison of 9-item vs 18-item categories, they showed increased false alarms when a single related target item had been presented (cf. Robinson & Roediger, 1997).

GENERAL DISCUSSION

The experiments reported here have provided answers to the two main questions we asked at the outset. First, our data indicate that both younger and older adults showed substantial false recognition of detailed color pictures that were categorically related to previously studied pictures, thereby extending the domain of false recognition phenomena. False recognition responses were particularly prominent for within-category lures that were drawn from categories that had a large number of exemplars at study (18 different exemplars) and so represented a strong and recurrent conceptual and/or perceptual ''theme'' during study. Across the three experiments, older adults falsely recognized 64% of these items; younger adults showed greater resistance to this type of error but still falsely recognized an average of 29% of these items—well above the baseline level of false alarms observed for novel-category lures (14 and 4% for older and younger adults, respectively). Note, however, that the levels of false recognition to lure pictures shown by younger adults, though substantial, are not nearly as high as the levels of false recognition to related lure words exhibited by younger adults in Roediger and McDermott's (1995) experiments. These observations are consistent with findings reported recently by Israel and Schacter (in press), who compared the false recognition of younger adults in the standard Deese/Roediger–McDermott paradigm where lists of associated words are studied, to a condition in which black and white line drawings corresponding to each word were also presented. Although significant levels of false recognition to related lures were documented following ''picture plus word'' encoding, the magnitude of the effect was considerably reduced compared to the standard ''word only'' encoding condition.

Second, older adults exhibited increased false recognition under conditions in which it is highly unlikely that they could have generated a lure at study and later exhibited source confusion about whether they had seen or only thought about the picture. Rather, older adults appeared to be vulnerable to false recognition because they were relying on general conceptual and/or perceptual similarity or gist-information that made it difficult to correctly identify as ''new'' items drawn from the categories that they had studied previously.

Our data also show that the processes that allow or support accurate recollection may be related to the processes that lead to errors in recollection and recognition. In all three experiments, older adults showed a much lower hit rate than younger adults for distinctive pictures that were not categorically related to other studied pictures and thus could not be as readily recognized on the basis of gistlike representations. Nonetheless, older adults showed a hit rate that closely approximated that shown by younger adults when the target items were drawn from categories for which they had seen many exemplars at study— items for which reliance on a gist representation might prove sufficient. Across the three experiments, the average hit rates for large-category targets were 81% for older adults and 82% for younger adults. However, for older adults, this was also accompanied by a considerable elevation in false recognition of lures drawn from those categories. Whereas older and younger adults showed essentially equivalent correct recognition rates for large-category items, the false recognition rates of older adults in each experiment doubled those of their younger counterparts (70 vs 35%, 58 vs 26%, and 63 vs 25% for Experiments 1, 2, and 3, respectively). Signal detection analyses examining hits compared to related false alarms revealed that younger adults had consistently greater sensitivity than older adults for items drawn from studied categories, but with only trends toward an overall age-related difference in response criteria (Experiments 2 and 3), or with older adults employing more lenient criteria, yet with both older and younger adults responding less conservatively to large-category items (Experiment 1).

While our experiments show consistently and unequivocally that older adults show greater false recognition of large-category pictures than do younger adults, there is the additional question of whether this is due to heightened sensitivity to gist (e.g., Gist-A′) or to more liberal responding to related lures from large categories. Considering only Experiments 2 and 3, where the categories were counterbalanced across category size, in Experiment 2, Gist-A′ for large categories for older adults was .83 compared to .73 for younger adults; in Experiment 3, the corresponding values were .77 for older and .71 for younger adults. In both cases, however, older adults also employed comparatively more lenient response criteria than did younger adults: $B_{D''}$ of .62 vs .88 in Experiment 2 and .24 vs .88 in Experiment 3.

Would older adults still show heightened sensitivity to gist if their level of false alarms to novel-category lures more nearly approximated that of the young? For Experiments 2 and 3, we attempted to address this question by selecting subsets of older and younger participants who were matched on their rate of false alarms to novel-category items. Specifically, we excluded any younger or older participants who never false alarmed to any novel-category items (false alarm rate of zero), and also any older adults who false alarmed more often than the highest false-alarming young adult. Application of this procedure to the data from Experiment 2 resulted in the exclusion of four older adults (two with false alarm rates of 0, and two with false-alarm rates exceeding 20%), and six younger adults (all with novel-category false-alarm rates of 0). This generated two subsets of older ($n = 12$) and younger ($n = 10$) participants, matched on their overall level of novel-category false alarms (averages of 9 and 8% for old and young, respectively, $F(1, 20) < 1$.

We then examined the pattern of hits and false alarms and the outcomes for the gist measures of sensitivity and response bias in these two subsets of matched older and younger adults. These analyses yielded three particularly pertinent outcomes:

(a) Older adults showed markedly increased false recognition of items from medium (.49) and large (.58) categories compared to younger adults (.26 and .33, respectively), $F(1,20) = 11.15$, $MS_e = .06$, $p = .003$, for the overall effect of age.

(b) For both medium (.80) and large (.84) categories, older adults showed higher levels of gist sensitivity (A′-Gist) than younger adults (.70 and .73, respectively), $F(1,20) = 8.62$, $MS_e = .01$, $p = .008$, for the overall effect of age.

(c) This difference in sensitivity was also accompanied by a trend toward a difference in response criteria, with older adults showing somewhat more lenient criteria for both medium- (.74) and large- (.65) category items than younger adults (.91 and .82, respectively), $F(1, 20) = 3.07$, $MS_e = .10$, $p = .10$, for the overall effect of age.

We also applied a similar procedure to the data from Experiment 3, excluding young subjects who never false alarmed to any novel category items ($n = 7$) and older adults who false alarmed more often than the highest false alarming young adult (i.e., any older adults with a novel false alarm rate greater than 20%; $n = 6$; no older adults in this experiment had zero false alarms to novel-category items). These exclusions resulted in two subsets of participants from Experiment 3 matched on baseline rates of false alarms: averages of 11 and 9% for old and young, respectively, $F(1, 17) < 1.2$. Examination of the data for these subsets of older and younger adults revealed the following patterns:

(a) Older adults' false recognition rates for within-category lures systematically increased with the number of related exemplars presented at study: older adults falsely recognized 24% of the items if they had previously encountered one related item at study, 45% if they had encountered 9 related items, and 60% if they had earlier been exposed to 18 similar items; the corresponding proportions for younger adults were 13% for single (1-item) lures, 36% for 9-item, and 27% for 18-item categories, $F(1,17) = 7.97$, $MS_e = .06$, $p = .01$, for the main effect of age.

(b) On the measure of Gist-A′, older adults showed greater sensitivity to large-category items (.81) than did younger adults (.67); this was also true for single-category items, though to a less pronounced extent (.64 and .57 for old and young, respectively), and was not seen for medium-category items (.76 vs .74 for old and young, respectively); $F(1,17) = 6.66$, $MS_e = .01$, $p = .02$ for the main effect of age with all three categories combined; $F(1,17) = 5.34$, $MS_e = .02$, $p = .03$ for large alone, $F < 1$ for medium alone, $F < 1.8$ for single alone.

(c) There was no overall difference in response criteria ($B_{D''}$) for older (.62) compared to younger (.67) adults, $F < 1$, but there was an interaction of age with category type, $F(2, 34) = 4.07$, $MS_e = .12$, $p = .03$. Subsequent pairwise comparisons indicated that older adults showed a trend toward greater leniency for items from large categories, $F(1,17) = 3.38$, $MS_e = .16$, $p = .08$ (averages of .47 vs .81 for old and young, respectively), no age difference in criteria for medium categories, $F < 1.5$, and a trend toward the reverse effect (older adults more conservative than younger adults) for single-category items, $F(1,17) = 2.81$, $MS_e = .14$, $p = .11$.

These additional analyses thus suggest that both differences in sensitivity to gist *and* differences in response criteria for items from studied categories may contribute to older adults' heightened likelihood of falsely recognizing within-category lures when a large number of exemplars were studied. Although the post hoc matching procedure is not without hazards, such as vulnerability to regression toward the mean effects, the generally similar patterns between the subgroup analyses for the two experiments provides some reassurance as to the reliability of these results: older adults show clear increased gist sensitivity for items for which a large number of related items were studied, together with a less pronounced tendency toward more lenient responding. Further studies might attempt to experimentally match older and younger adults on the baseline measure of false alarms, so as to further assess the relative roles of increased sensitivity to gist, vs differences in response criteria, in the behavioral phenomena that we have demonstrated.

These observations are related to Brainerd and Reyna's fuzzy trace theory (Brainerd et al., 1995; Reyna & Brainerd, 1995), which holds that correct positive responses to targets on a recognition test are most often driven by an item-specific ''verbatim'' trace that preserves surface details of the stimulus whereas false alarms are largely driven by an independent gist trace based on memory for the senses and meanings that were accessed during initial encoding. Our results with young adults are, in some respects, consistent with this idea: In all three experiments, the hit rates (and measures of sensitivity) of younger adults were relatively invariant across the different item types (unrelated, single, small, medium, and large) whereas false alarms generally increased with category size (comparing novel false alarms to related false alarms, though not consistently in comparing false alarms for medium vs large categories). However, our data concerning older adults indicate that hits, like false alarms, may depend heavily on gist or general similarity information. In light of these data, it is possible that hits to categorized pictures for younger adults, too, were influenced by gist representations. However, unlike older adults, younger participants could also draw on robust, distinctive item-specific representations when gist representations were less likely to be highly accessible (i.e., with unrelated pictures). In other words, it is possible that, although younger adults' performance for the categorized and unrelated target items was behaviorally indistinguishable, the underlying cognitive representations or processes that led them to give a positive recognition response to categorized targets differed from those that led them to give positive recognition responses to unrelated items. In this connection, Hintzman (1988) has pointed out that several theories of recognition memory predict that *both* correct and false recognition of probes should increase with the number of list items that partially match the probe; moreover, using categorized and noncategorized words as stimuli, he found that young subjects (undergraduates) showed a small but significant linear trend in hits, such that hits increased monotonically with the number of same-category items in the list. Future studies will be needed to determine the extent to which, and conditions under which, hit rates to categorized pictures in younger adults might also be characterized by substantial reliance on generic or gistlike representations.

The conjunction of a high level of hits and a high level of false alarms that we observed for older adults with pictorial stimuli has been reported previously for elderly individuals in the verbal domain. For example, Reder, Wible, and Martin (1986) reported that older adults showed hit rates equivalent to or slightly exceeding those shown by younger adults when the target items were statements that could plausibly have been incorporated in short stories they had read and that were, in fact, actually presented. However, older adults showed particularly poor performance for statements that, though plausible, were not actually presented—often falsely recognizing these items. These researchers attributed this age difference to older adults' adoption of a ''plausibility'' rather than ''direct retrieval'' strategy, observing that when the plausibility strategy was used during recognition testing, previously presented (and plausible) facts would often be judged accurately yet not-presented (but also plausible) statements would tend to be erroneously accepted. They proposed that older adults were relying on a relatively automatic ‘‘partial matching’’ strategy in making their plausibility judgments, wherein rather than ‘‘carefully matching’’ the verbal statements against specific propositions in memory, they relied on a process of detecting feature overlap between the test items and information in memory (also see Bartlett, Leslie, Tubbs, & Fulton, 1989).

This account focuses on strategies that the elderly may have employed at the time of retrieval and that may have rendered them vulnerable to within-category or gist-based errors and is similar to distinctions between familiarity- or similarity-based processes that may support recognition, and retrieval of more specific content (e.g., Mandler, 1980) that may be used to suppress or oppose possibly misleading familiarity information (e.g., Hintzman & Curran, 1994, 1995; Jacoby, 1991). However, the age-related elevation of within-category false recognition responses found in the three experiments reported here may also have involved differential processes or strategies operative at encoding or an interaction of encoding and retrieval factors. For example, it is possible that the encoding processes of older adults involved a greater reliance on the generation of the category labels of the stimuli at study (e.g., ‘‘chair,’’ ‘‘teddy bear’’) than did that of younger adults, who may have focused more on perceptual details of individual items. To some extent, the use of an orienting task may have mitigated against such differences at encoding—all participants were required to provide ‘‘liking’’ ratings of the stimuli and were encouraged to base their ratings on each stimulus individually rather than responding to the general class to which the item belonged. Nonetheless, it is possible that the names of the items assumed a larger role in the encoding processes of the old than those of the young, possibly because access to this form of semantic information is comparatively effortless or automatic (cf. Intraub & Nicklos, 1985; Smith & Magee, 1980) or because older adults (the instructions to the contrary notwithstanding) more often made their judgments of liking on the basis of the object categories rather than the specific exemplars they were shown. To the extent that the label provides only conceptual information that would be of comparatively little use in differentiating studied from nonstudied exemplars of the same category, the elderly might be disadvantaged either because their reliance on this conceptual information would lead them to incorrectly embrace as ‘‘old’’ items that were, in fact, ‘‘new,’’ or because this conceptual information simply did not provide grounds for the rejection of within-category lures. Not inconsistent with this possibility is evidence reported by Marks (1991, Experiment 4) from a comparison of different encoding tasks used with line drawings. Undergraduates who answered distinct orienting questions that focused on *categorical* information concerning an object (e.g., is it a mode of travel? a type of vehicle? a form of transportation?) showed higher rates of false alarms than did participants who were asked orienting questions that focused on physical features of the object (is it red? is it round?) or who were asked if an object would ‘‘fit’’ in a visually presented scene (though this effect was observed only when the answer to the orienting question was affirmative).

More generally, there is evidence that the amount of pictorial detail that is retained depends on characteristics of processing at both encoding and test (e.g., Pezdek et al., 1988) and that picture recognition improves as more information about visual details is noted and encoded (e.g., Intraub & Nicklos, 1985; Loftus & Kallman, 1979; Potter, 1976). Loftus and Bell (1975) found that when subjects were able to indicate that they were making a recognition decision on the basis of a particular detail in a line drawing or photograph, rather than simply responding on the basis of the familiarity of the picture, discriminability (d') was, on average, increased by 1.5. They used the terms ‘‘specific detail’’ and ‘‘general visual’’ information to refer to these two classes of information, respectively. Similarly, Loftus and Kallman (1979) found that participants who were encouraged, at time of study, to identify specific details that would help them recognize complex pictures later showed enhanced ability to discriminate pictures relative to those who did not name details during acquisition. To the extent that younger adults in the present experiments were spontaneously more likely to notice specific physical features of the stimuli whereas older adults were less likely to do so, older adults might be expected to show enhanced susceptibility to false recognition due to items being congruent with the general classes of things they had studied. Future experiments using categorized pictures such as those used in the present studies could modify the encoding conditions so as to increase attentional focus on critical perceptual information and also to increase careful examination or scrutiny of items at test.

Such experiments will provide further information regarding the exact conditions under which heightened susceptibility to false recognition may be found in both older and younger adults. Nonetheless, the present experiments clearly demonstrate that accounts of gist-based false recognition cannot exclusively focus on deficits surrounding susceptibility to source confusions. It is highly unlikely that individuals, during the study phase of our experiments, spontaneously generated pictorial representations of the exemplars that were later lures on the recognition test. Rather, age differences in how specifically or generally the items were encoded (Isingrini et al., 1995; Rabinowitz & Ackerman, 1982; Rabinowitz et al., 1982; Schacter et al., in press) or how carefully individuals scrutinized generally familiar or similar-seeming materials before designating them as ''old'' (Johnson et al., 1993; Mather, Henkel, & Johnson, in press; Multhaup, 1995; Schacter et al., in press; also cf. Dodson & Johnson, 1993; Gauld & Stephenson, 1967; Hintzman & Curran, 1994, 1995) become the focus of interest. It is worth emphasizing in this regard that, although our data are not readily accommodated by a source confusion account in the strict sense in which we have used the notion of ''source confusion,'' our results may nonetheless reflect a source monitoring deficit in a broader sense (Johnson et al., 1993). For instance, recent evidence from brain imaging research indicates that, under conditions of difficult retrieval, elderly individuals do not activate anterior regions of the frontal lobes that show robust activation during effortful search in younger adults (Schacter, Savage, Alpert, Rauch, & Albert, 1996). If elderly adults failed to engage in the effortful retrieval or monitoring activities needed to distinguish studied pictures from categorically related lures, increased false recognition would be expected (cf. Curran et al., 1997; Schacter, Curran, Galluccio, Milberg, & Bates, 1996). This kind of source monitoring deficit (Johnson et al., 1993) may interact with, or be the result of, older adults' initial failure to encode distinctive properties of target pictures. Extending the type of stimuli employed in research on memory inaccuracy to include richly detailed pictorial items such as those employed in the present study should help to provide a more complete account of processes underlying inaccuracy in both older and younger adults. Equally important, the incorporation of appropriate comparison conditions of studied items, including both categorized and unrelated items, may further inform us as to the extent to which apparently accurate memory—correct recognition—is supported by precisely the same forms of information that underlie inaccurate memory.

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.

Anisfeld, M., & Knapp, M. (1968). Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology,* **77,** 171–179.

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

Bousfield, W. A., Whitmarsh, G. A., & Danick, J. J. (1958). Partial response identities in verbal generalization. *Psychological Reports,* **4,** 703–713.

Bower, G. H., & Glass, A. L. (1976). Structural units and the redintegrative power of picture fragments. *Jour-*

nal 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.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology,* **2,** 331–350.

Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin,* **90,** 513–546.

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

Craik, F. I. M. (1977). Age differences in human memory. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 384–420). New York: Van Nostrand Reinhold.

Craik, F. I. M., Anderson, N. D., Kerr, S. A., & Li, K. Z. H. (1995). Memory changes in normal ageing. In A. D. Baddeley, B. A. Wilson, & F. N. Watts (Eds.), *Handbook of memory disorders* (pp. 211–241). New York: Wiley.

Craik, F. I. M., & Rabinowitz, J. C. (1985). Age differences in the acquisition and use of verbal information: A tutorial review. In H. Bouma & 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.

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.

Frieske, D. A., & Park, D. C. (1993). Effects of organization and working memory on age differences in memory for scene information. *Experimental Aging Research,* **19,** 321–332.

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

Goldstein, A. G., & Chance, J. E. (1970). Visual recognition memory for complex configurations. *Perception and Psychophysics,* **9,** 237–241.

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. (1989). Aging and source monitoring. *Psychology and Aging,* **4,** 106–112.

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

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 and Cognition,* **23,** 213–226.

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

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **18,** 180–191.

Intraub, H., & Nicklos, S. (1985). Levels of processing and picture memory: The physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **11,** 284–298.

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. (in press). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin and Review.*

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language,* **30,** 513–541.

Johnson, M. K., Bransford, J. D., & Solomon, S. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology,* **98,** 203–205.

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

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.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review,* **87,** 252–271.

Marks, W. (1991). Effects of encoding the perceptual features of pictures on memory. *Journal of Experi-*

mental Psychology: Learning, Memory, and Cognition, **17,** 566–577.

Mather, M., Henkel, L. A., & Johnson, M. K. (in press). Evaluating characteristics of false memories: Remember/know judgments and Memory Characteristics Questionnaire compared. *Memory and Cognition.*

McIntyre, J. S., & Craik, F. I. M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology,* **41,** 175–192.

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

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology,* **19,** 155–160.

Norman, K. A., & Schacter, D. L. (in press). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory and Cognition.*

Park, D. C., & Whitten, W. B. (1977). Abstraction of linguistic, imaginal, and pictorial ideas. *Journal of Experimental Psychology: Human Learning and Memory,* **3,** 525–538.

Paul, L. M. (1979). Two models of recognition memory: A test. *Journal of Experimental Psychology: Human Learning and Memory,* **5,** 45–51.

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.

Pezdek, K., Maki, R., Valencia-Laver, D., Whetstone, T., Stoeckert, J., & Dougherty, T. (1988). Picture memory: Recognizing added and deleted details. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 468–476.

Phelps, E. A., & Gazzaniga, M. S. (1992). Hemispheric differences in mnemonic processing: The effects of left hemisphere interpretation. *Neuropsychologia,* **30,** 293–297.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science,* **1,** 125–126.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory,* **2,** 509–522.

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.

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 and Review,* **3,** 105–111.

Reder, L. M., Wible, C., & Martin, J. (1986). Differential memory changes with age: Exact retrieval vs 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. (1997). Associative processes in false recall and recognition. *Psychological Science,* **8,** 231–237.

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

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 803–814.

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* (pp. 1–43). Cambridge, MA: Harvard Univ. Press.

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

Schacter, D. L., Kaszniak, A. K., Kihlstrom, J. F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging,* **6,** 559–568.

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. (in press). The cognitive neuroscience of constructive memory. *Annual Review of Psychology.*

Schacter, D. L., Savage, C. R., Alpert, N. M., Rauch, S. L., & Albert, M. S. (1996). The role of hippocampus and frontal cortex in age-related memory changes: A PET study. *NeuroReport,* **7,** 1165–1169.

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.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior,* **6,** 156–163.

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.

Smith, A. D. (1975). Partial learning and recognition memory in the aged. *International Journal of Aging and Human Development,* **6,** 359–365.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General,* **109,** 373–392.

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.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology,* **25,** 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science,* **19,** 73–74.

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

Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology,* **103,** 255–262.

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1996). *Older adults show greater susceptibility to false memory than young adults: Temporal characteristics of false recognition.* [Manuscript submitted for publication]

Tussing, A. A., & Greene, R. L. (in press). False recognition of associates: How robust is the effect? *Psychonomic Bulletin and Review.*

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

Vogt, J., & Kimble, G. A. (1973). False recognition as a function of associative proximity. *Journal of Experimental Psychology,* **99,** 143–145.

Wyant, S., Banks, W. P., Berger, D., & Wright, P. W. (1972). Verbal and pictorial similarity in recognition of pictures. *Perception and Psychophysics,* **12,** 151–153.

(Received January 9, 1997)
(Revision received June 24, 1997)