Journal of Experimental Psychology:
Learning, Memory, and Cognition
2003, Vol. 29, No. 4, 499–510

0278-7393/03/$12.00 DOI: 10.1037/0278-7393.29.4.499

False Recognition of Abstract Versus Common Objects in Older and Younger Adults: Testing the Semantic Categorization Account

Wilma Koutstaal
Harvard University and University of Reading

Chandan Reddy, Eric M. Jackson, Steve Prince,
Daniel L. Cendan, and Daniel L. Schacter
Harvard University

Older adults often demonstrate higher levels of false recognition than do younger adults. However, in experiments using novel shapes without preexisting semantic representations, this age-related elevation in false recognition was found to be greatly attenuated. Two experiments tested a semantic categorization account of these findings, examining whether older adults show especially heightened false recognition if the stimuli have preexisting semantic representations, such that semantic category information attenuates or truncates the encoding or retrieval of item-specific perceptual information. In Experiment 1, ambiguous shapes were presented with or without disambiguating semantic labels. Older adults showed higher false recognition when labels were present but not when labels were never presented. In Experiment 2, older adults showed higher false recognition for concrete but not abstract objects. The semantic categorization account was supported.

Age-related reductions in the ability to intentionally remember recently experienced events are well documented for a wide variety of types of stimuli (for reviews, see Anderson & Craik, 2000; Balota, Dolan, & Duchek, 2000). However, in addition to these well-known, age-related reductions in memory for recent experiences, as demonstrated by an increased number of misses or errors of omission in older adults, a growing number of studies show that *false positive* errors—those that involve false recognition, false recall, or errors of commission, in which new items are misidentified as old—may also be more likely in older than in younger adults (e.g., Balota et al., 1999; LaVoie & Faulkner, 2000; Norman & Schacter, 1997; Searcy, Bartlett, & Memon, 1999; Smith, 1975; Tun, Wingfield, Rosen, & Blanchard, 1998; for reviews, see Schacter, Koutstaal, & Norman, 1997; Schacter, Norman, & Koutstaal, 1998). Age-related differences in false recognition have been found for both semantically related words and semantically related pictures. Age differences have been found to be especially marked for categorically related pictures of common objects (Koutstaal & Schacter, 1997; Koutstaal, Schacter, & Brenner, 2001; Koutstaal, Schacter, Galluccio, & Stofer, 1999). For these types of stimuli, false recognition rates of older adults have reached as high as 60%–70%, compared with 25%–30% in younger adults.

In the experiments mentioned above, the lure items were both conceptually and perceptually similar to the studied items (pictures of common objects). The question addressed in the present experiments was whether similar outcomes would be observed if the previously encountered items and the lures share only perceptual similarity and do not belong to preexisting semantic or conceptual categories. One possible interpretation of the age-related increase in false recognition responses for common objects (and the more modest increases for semantically related words) is that in situations in which both perceptual and conceptual information is present, older adults may primarily focus on the semantic or conceptual information. This might arise because semantic information is more readily or automatically accessed or for other reasons (cf. Brainerd & Reyna, 1998; Koutstaal & Schacter, 1997; Schacter et al., 1998; see also Hess, 1990, for a more general discussion of possible age-related differences with emphasis on conceptually-driven vs. data-driven processing). An emphasis on semantic information (e.g., the names of objects) could tend to exacerbate false recognition responses by underscoring the similarity between exemplars and by attenuating or detracting from the encoding of differentiating perceptual features. We refer to this simple idea as the *semantic categorization account*: Preexisting semantic or conceptual information detracts from the processing of nonconceptual, item-specific perceptual information. Note that this interpretation does not imply that older adults encode semantic information more deeply. Indeed, there is good reason and evidence to suppose that most often older adults do not semantically elaborate on materials, with consequent costs to later recall and memory (e.g., Craik, 1982). Rather, the key notion is that semantic

Wilma Koutstaal, Department of Psychology, Harvard University, and Department of Psychology, University of Reading, Reading, United Kingdom; Chandan Reddy, Eric M. Jackson, Steve Prince, Daniel L. Cendan, and Daniel L. Schacter, Department of Psychology, Harvard University.

The data reported in Experiment 1 formed part of the Honors Bachelor of Arts thesis of Chandan Reddy, which was presented to the Department of Psychology and the Concentration in Biology, Harvard University, Cambridge, Massachusetts, March 2000. This research was supported by National Institute on Aging Grant AG08441.

We thank an anonymous reviewer for helpful comments. Also, we thank Daniel Cousins for help with modification of the stimuli used in one of the preliminary experiments, Alison Carter and Jean Rhoo for creation of the stimuli used in Experiment 2, and Aoife Fitzgerald for help in scanning in the stimuli and collecting normative data in Experiment 2.

Correspondence concerning this article should be addressed to Wilma Koutstaal, Department of Psychology, University of Reading, Earley Gate, Reading RG6 6AL, United Kingdom. E-mail: w.koutstaal@reading.ac.uk

category information truncates, precludes, or preempts further item-specific processing, even though the initial categorization is quite straightforward and effortless. Evidence consistent with such a possible effect in younger adults who were given conceptually oriented encoding instructions was provided in a study by Marks (1991, Experiment 4; see also Intraub & Nicklos, 1985; Loftus & Kallman, 1979).

Another possible interpretation, however, is that older adults generally demonstrate reduced encoding of differentiating perceptual details that would allow successful discrimination of studied items from similar-appearing items, irrespective of the copresence of semantic information. According to this account, whereas younger adults may efficiently encode and later use the comparatively rich and detailed perceptual information provided by visual stimuli (whether concrete or abstract), this information may be less often, or less efficiently, encoded and subsequently used by older adults (e.g., Park, Puglisi, & Sovacool, 1983, 1984; Trahan, Larrabee, & Levin, 1986; Winograd, Smith, & Simon, 1982; but also cf. Park, Puglisi, & Smith, 1986), regardless of the nature of the stimuli. We refer to this idea as the *impaired perceptual encoding hypothesis.*

Although a number of studies with younger adults have examined false recognition of perceptually similar items in which the materials are not readily named or semantically categorized (e.g., Bower & Glass, 1976; Solso & Raynis, 1979), few have considered older adults (but see Hess, 1982). In a series of previous experiments (e.g., Jackson, 1998), we examined false recognition of older and younger adults for novel abstract stimuli that did not belong to any preexisting semantic or conceptual categories (stimuli as used in Koutstaal, Schacter, Verfaellie, Brenner, & Jackson, 1999). In these experiments, we found that the two age groups showed only modest, and nonsignificant, differences in overall false recognition (irrespective of confidence level), although a modest and significant age difference did emerge when considering only highly confident responses. The findings from one of those experiments are shown in Figure 1. These experiments provide partial support for the semantic categorization account.

However, although suggestive that semantic information may exacerbate age differences in false recognition responding, these earlier experiments did not directly compare false recognition for semantically versus nonsemantically related materials. Thus, they did not directly address the key question of whether it is particu-

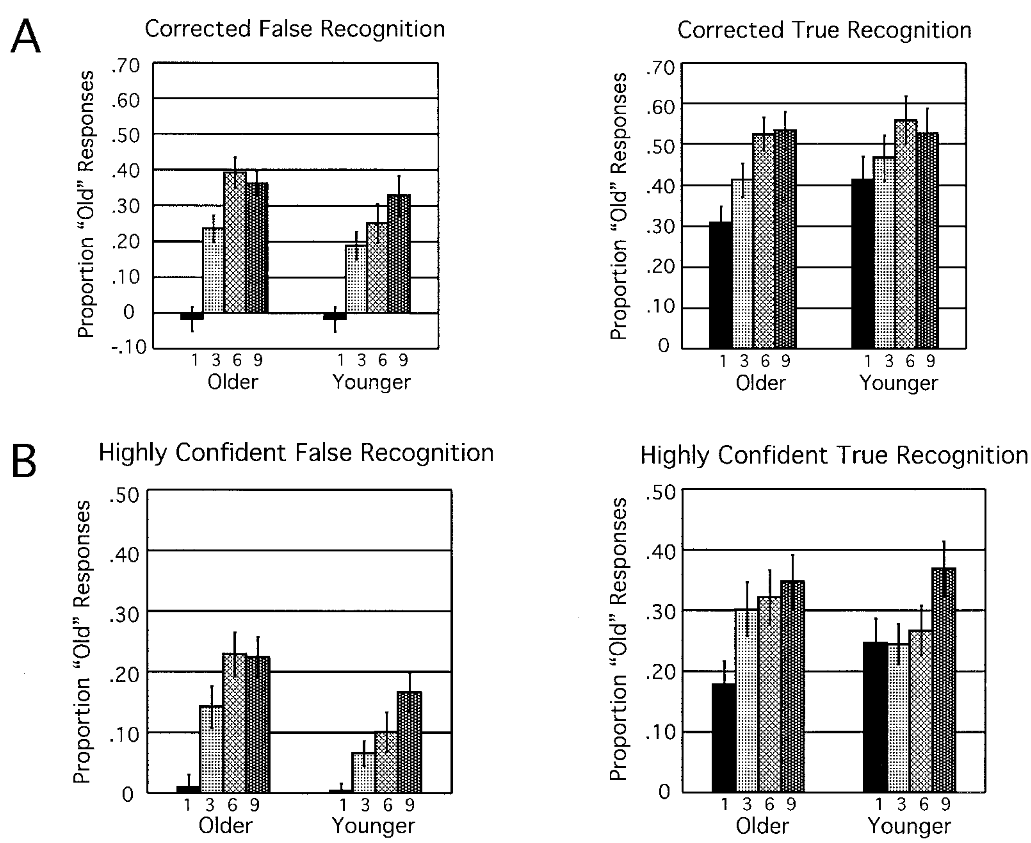

Figure 1. A: Mean proportions of "old" responses to nonstudied items (false recognition; left) and studied items (true recognition; right) after subtracting false alarms to novel category items, for one of the preliminary experiments (Koutstaal, 2000) involving abstract figures. Results are shown separately as a function of age (older adults or younger adults) and category size (one, three, six, or nine categorically related exemplars presented at study). B: Mean proportions of highly confident "old" responses for the same conditions shown in A. Results are shown after subtracting highly confident false alarms to novel category items. Error bars represent standard errors of the means.

larly for stimuli involving preexisting semantic representations that older adults show increased false recognition. The two experiments reported here were designed to address this question. In Experiment 1, we used a novel experimental approach involving ambiguous stimuli to evaluate the effect of the presence versus the absence of semantic information on false recognition while holding the perceptual stimulus constant. Experiment 2 provided a within-experiment and within-subject test of differences in false recognition for novel versus common objects for older compared with younger individuals.

Experiment 1

In this experiment, we adopted a novel approach that allowed the analytical examination of false positive responding under conditions in which the presence or absence of semantic information was manipulated during the initial encoding of an object and/or during attempted recognition judgments (at retrieval). The approach involved the creation of comparatively simple but ambiguous visual figures or objects. (See Figure 2 for examples.) When presented alone (without a disambiguating label), these figures were not consistently identified as belonging to any one common object category. However, if a label was provided, the figures were such that older and younger participants most often agreed that the depicted object could belong to the category indicated. In the experimental study, we then simply manipulated whether the disambiguating label was presented together with the figures during the initial presentation of the stimuli (at encoding) and/or during testing (at retrieval). Four conditions were created by the orthogonal combination of whether verbal labels of the stimuli were presented during encoding and/or retrieval: *clear–clear* (CC; no label at either encoding or retrieval), *label–label* (LL; label at both encoding and retrieval), *label–clear* (LC; label at encoding but no label at retrieval), and *clear–label* (CL; no label at encoding but label at retrieval).

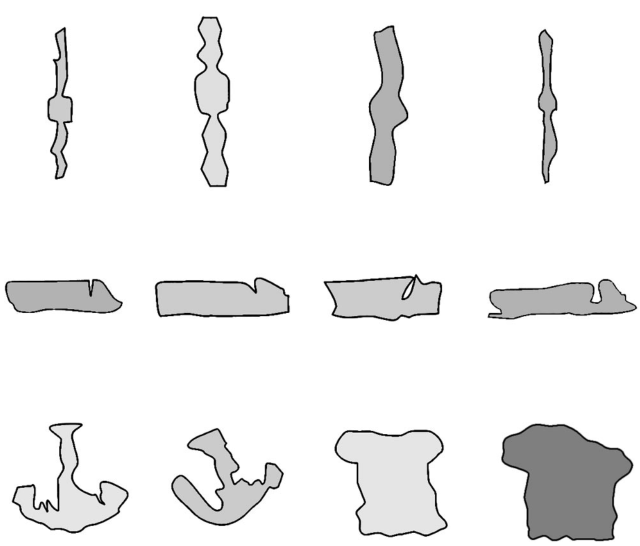

Figure 2. Examples of the ambiguous stimuli used in Experiment 1. Top and middle: Exemplars of the categories *watch* and *truck,* respectively. Bottom: Unrelated pairs for *anchor* and *bread.* The first three watches and trucks are critical items, whereas the fourth watch and truck are noncritical items. Note that although, for illustrative purposes here, the stimuli are shown in black and white, the stimuli as shown to participants were presented in color, with color being an important attribute that was varied both within and across categories.

If the semantic categorization account is correct, then an age-related elevation in false recognition responses should be found only in the conditions in which the semantic labels are presented with the figures. In the most straightforward instance, this idea leads to the prediction that there will be an age-related increment in false recognition responses in the LL condition but no age-related difference in false recognition responses in the CC condition. The predictions for the two mixed conditions, those involving labels present during encoding but not during recognition testing (LC) or not during encoding but during recognition testing (CL), are less straightforward. However, comparisons of these conditions with the LL and CC conditions should provide an indication of the relative importance of conceptual information during initial processing versus attempted retrieval in determining levels of false recognition. In contrast, if the impaired perceptual encoding account is correct, then an age-related elevation of false recognition responses should be observed in all conditions and, in the strongest instance, in the CC condition alone, in which no labels were presented.

We also manipulated the number of categorically related items that were presented during the initial study phase. On the basis of previous studies (e.g., Koutstaal & Schacter, 1997; Robinson & Roediger, 1997), we expected that categories for which many similar items were presented at study would allow the extraction and retention of a robust representation of the perceptual or conceptual gist of the items and therefore produce higher levels of false recognition responding. In contrast, there should be a comparatively weaker representation of the gist of the items from categories for which no other related items were presented, with correspondingly lower levels of false recognition.

Method

Participants. A total of 72 older adults and 72 younger adults took part in the experiment. Older individuals were recruited from the community through newspaper advertisements and posted flyers and were screened for various neuropsychological and medical conditions that might interfere with performance, including a history of alcoholism or substance abuse, present or previous treatment for psychiatric illness, current treatment with psychoactive medication, drug toxicity, and primary degenerative brain disorders and brain damage sustained earlier from a known cause. Older participants were also given a screening questionnaire regarding the optimal time of day for them and were tested within their optimal times. Younger adults were Harvard University undergraduates recruited through sign-up sheets posted at the university. Both older and younger adults were screened for depression. All participants had normal or corrected-to-normal vision and were paid for their involvement in the study. The data from 1 additional younger and 4 additional elderly individuals who did not follow experimental instructions were eliminated.

The average age of older participants was 69.4 years (SD = 3.93; range = 60–75 years); they reported an average of 16.0 years (SD = 2.96) of formal education. The average age of the younger participants was 19.8 years (SD = 2.53; range = 17–30 years); they had on average 13.0 years (SD = 1.40) of formal education. (Educational data were not available for 4 younger participants.) Older adults had significantly more years of education than did younger adults, $F(1, 132) = 56.88$, $MSE = 5.33$, $p <$

.0001, but education level did not vary by experimental condition, and there was no Age × Condition interaction (Fs < 1.85).

Design. The experimental design included two between-subjects variables: age (older vs. younger) and labeling condition (CC, LL, LC, and CL). In addition, the design included a within-subjects factor of category size. For studied items (targets), category size had three levels: single (target items for which only one categorically related item had been presented during study), large (target items for which six categorically related items had been presented during study), and unrelated (noncategorized target items, unrelated to other items at study or test). For nonstudied items, category size had four levels: single (related lure items for which one categorically related item had been presented at study), large (related lure items for which six categorically related items had been presented at study), unrelated (unrelated lures that were unrelated to any items at study or test), and novel (lure items for which no categorically related items had been presented during study but from categories that across participants were systematically counterbalanced across study and test status). The novel category items provided an index of the baseline level of false alarms to categorized items.

Stimulus materials. The stimuli were ambiguous pictorial depictions of common objects, created using a computerized graphics program (Aldus FreeHand 7.0; Macromedia, 1996) and then converted to an appropriate format for computer screen display during testing. Selection of the stimuli involved an iterative procedure of stimulus creation, collection of normative ratings, modification of stimuli, and further collection of normative data. Normative ratings were collected and analyzed for three separate dimensions. These were (a) *naming*: evaluation of how often the objects were named under conditions in which no labels were presented, used to establish suitable levels of ambiguity of the stimulus items; (b) *label–picture agreement*: evaluation of the extent to which participants agreed with the labels of the objects when shown an object in combination with its intended name, used to establish that the stimuli, although ambiguous, could nonetheless indeed be viewed as belonging to the intended categories when this was suggested; (c) *shape sorting*: assessment of the extent to which participants sorted the stimuli, entirely on the basis of their visual form, into unique nonoverlapping categories corresponding to those intended during creation of the stimuli. A detailed description of the norming procedure is available in Reddy (2000). In brief, the successive stages aimed to create a set of stimuli that, although not often spontaneously named as belonging to a given category, would nonetheless be well agreed as belonging to a specified category under circumstances in which the name was provided and in which the task was to indicate whether the item could belong to that category. The third set of normative ratings involving the shape sorting task was used to ensure the independence of the stimulus categories. These ratings were used to identify categories of objects that were perceived as perceptually similar to one another; categories that showed modest to high levels of shape-based cross-category confusions were removed from the stimulus pool.

This iterative norming procedure, conducted with both older (n = 19) and younger (n = 19) adults, none of whom participated in the subsequent experimental sessions, yielded a total of 18 object categories, with each category comprising 9 exemplars. For counterbalancing purposes, these 18 categories were assigned to three sets of six categories each (to be counterbalanced across the single, large, and novel category size conditions). Examples of the items are provided in Figure 2, which also presents a few instances of unrelated items. (The unrelated items were included to increase the length and variety of the study and test lists, to avoid possible ceiling effects, and to increase comparability with previous experiments. For brevity, analyses of the unrelated items are not reported.)

To avoid confounding the size of the category at study with the size of the category at test, we tested only a subset of the studied items, termed *critical items,* during recognition, thus holding category size at test constant while manipulating category size only during encoding. Specifically, with one exception (which we describe later), each category was tested three times with studied items (targets) and three times with categorically related lures. To accomplish this, we selected six items from each category to serve as critical items. These were assigned to two equal subsets of three items each (Critical A and Critical B), and across participants, these subsets were rotated across study–test status. Thus, items from the two subsets were tested for all participants, but whereas Critical A comprised targets and Critical B comprised lures for some participants, the reverse held for other participants (Critical A comprised lures and Critical B comprised targets). The critical items were selected on the basis of the normative rating data as the best exemplars of each category—that is, items that for both older and younger adults showed a level of ambiguity as high as possible in conjunction with high levels of label–picture agreement and minimal rates of across-category shape classifications. The critical subsets were equated on these ratings. The one exception concerned single items (i.e., items for which only one item in its category was shown at study). For these items, each category was tested once with that studied item (target) and once with a related lure. One Critical A and one Critical B item for each category was selected for this purpose; these items were chosen to be intermediate in the normative ratings of the critical items and were counterbalanced across studied and nonstudied status.

Procedure. The overall procedure involved three phases, including a study phase in which participants were exposed to the stimuli under an incidental encoding task, a brief retention interval, and the test phase. All participants were tested individually in a single experimental session of approximately 45–60 min. The stimuli at both study and test were presented using a color computer monitor and PsyScope experimental presentation software (Cohen, MacWhinney, Flatt, & Provost, 1993). The stimuli appeared in the center of the screen, with prompts for responding to the encoding task or the recognition task displayed beneath.

In the study phase, each item was presented for 3 s, and participants were asked to rate (using the number keys of the computer keyboard) how much they liked the visual appearance of the stimulus (1 = *do not like*; 5 = *very much like*). This encoding task was chosen as one that was likely to encourage attentive processing of the individual stimulus items and as similarly applicable for all of the labeling conditions.

During the study phase, for participants in the LL condition and in the LC condition, the label of each stimulus appeared for 2 s prior to the presentation of the picture. For clear conditions in which a label was not presented, participants were shown a fixation crosshair for 2 s prior to the presentation of the picture. To ensure that participants attended to the labels, for all conditions involving labels (both during study and during testing), we required participants to read the labels out loud. Each participant was exposed to a total of 60 items at study, including 6 items from each of 6 large categories, 1 item from each of 6 single categories, 12 unrelated items, and 3 primacy and 3 recency buffers. The items during study were presented in a pseudorandom order such that items from the different category sizes were distributed equally throughout the list. The study phase was followed by a 10-min retention interval, during which participants performed unrelated filler tasks.

In the test phase, participants were shown a subset of the items that had been presented in the study phase, together with new items (both related lures and novel items). Participants were asked to indicate, using designated keys of the computer keyboard, whether each item was old (i.e., had been presented during the prior visual-appearance-liking rating task) or new (i.e., had not previously been presented in the experiment) and then to indicate their degree of confidence in their recognition judgments (1 = *just guessing*; 5 = *very sure*). Each test list consisted of a total of 90 items: 18 large category targets (3 items × 6 categories), 18 large category lures (3 items × 6 categories), 6 single category targets (1 item × 6 categories), 6 single category lures (1 item × 6 categories), 18 novel category lures (3 items × 6 categories), 12 unrelated targets, and 12 unrelated lures. New and old test items were presented in a pseudorandom order such that items from the different category sizes were evenly distributed across the test.

Results

The outcomes of the recognition test are shown in Figure 3, which presents the proportion of "old" responses for lure items (false recognition) and studied items (true recognition) separately as a function of age, labeling condition, and category size (including novel items). From Figure 3, it is clear that, as expected, higher levels of false recognition were observed for the large categories. Although this pattern was found for both older and younger adults, it appears to be especially marked for older adults in the LL condition (and, to a lesser degree, in the LC condition). For true recognition, the two age groups show similar levels of recognition for the large categories, but older adults show impaired recognition for the single items. Figure 4 presents the results after correction for false alarms to the novel category lures, which provided a baseline level of responding "old" to new items. For ease of graphic presentation and for reasons detailed below, Figure 4 focuses on the two extreme conditions, for which strongest predictions can be made (LL and CC).

False recognition. We first assessed whether there were systematic effects of age or labeling condition on the baseline level of false alarms to novel category items. A 2 (age) × 2 (label at study vs. no label at study) × 2 (label at test vs. no label at test) between-groups analysis of variance (ANOVA) revealed no overall effect of age ($F < 1$). Combining across all conditions, older and younger adults showed identical rates of false positive errors to novel category items ($M = .12$ for both age groups). There was, however, a highly significant effect of label at study, $F(1, 136) = 14.54$, $MSE = 0.01$, $p = .0002$. Baseline rates of false positive responding were twice as high in the conditions in which there was no label presented during study (i.e., under clear conditions at study, $M = .16$) than under conditions in which a label was shown at study ($M = .08$). Although this pattern was shown by both age groups, it was somewhat more pronounced for the younger adults (a decrease in novel false alarms of 4% for the older adults compared with 11% for the younger adults), $F(1, 136) = 2.86$, $MSE = 0.01$, $p = .09$, for the Age × Label at Study interaction. There were no other effects (Fs < 1).

This effect of labeling condition on the baseline rate of false alarms somewhat complicates analyses for categorized items, because labeling condition effects following correction for baseline false alarms might then either simply reflect this initial difference or reflect this difference plus further differences in categorized item responding. Moreover, this effect on the baseline rates is in the opposite direction to that expected within the semantic categorization account for the categorized items in the older adults. According to that account, older adults should show elevated false recognition of the category items when the label is present (not absent). For this reason, we present the analyses for categorized items both for the raw responses and following correction for false alarms to novel category items (referred to as *novel-corrected*). However, given these additional analyses, in the interests of brevity we focus on the two conditions in which the strongest predictions could be made, that is, the LL and CC conditions.

Consistent with the semantic categorization account, a 2 (age: older or younger) × 2 (category size: large or single) × 2 (condition: LL or CC) analysis on the raw false recognition measure showed a significant Age × Labeling Condition interaction, $F(1, 68) = 4.88$, $MSE = 0.05$, $p = .03$. This interaction arose because there was a greater increment in false recognition in older adults than in younger adults for the LL compared with the CC condition (an increment of 12% for the older adults compared with a decrement of 5% for the younger adults). This analysis also showed a main effect of age, with older adults showing a higher overall level of false recognition than younger adults, $F(1, 68) = 7.52$,

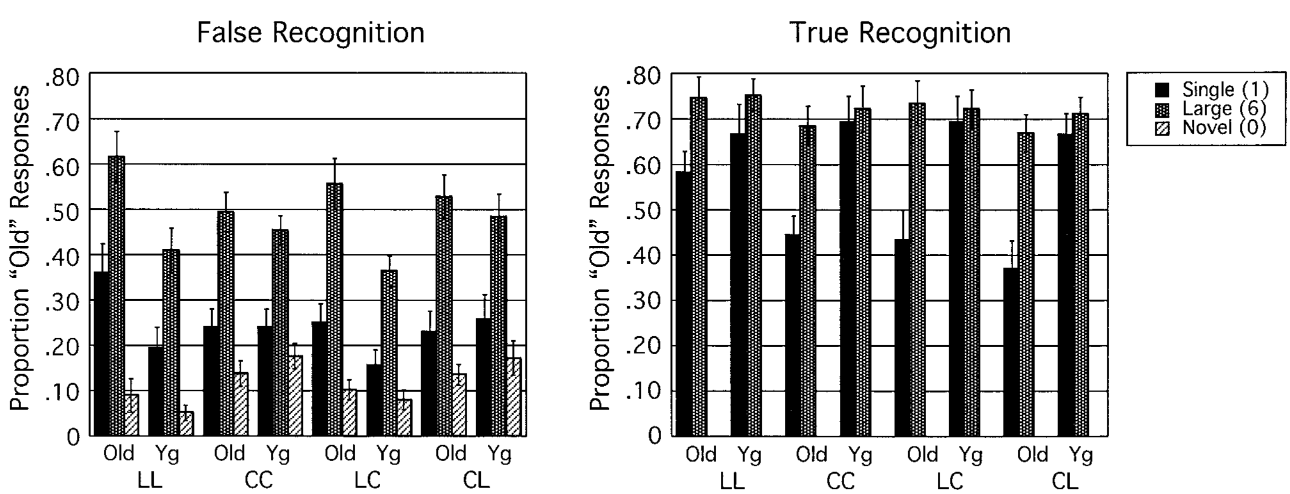

Figure 3. Proportions of "old" responses to nonstudied items (false recognition; left) and studied items (true recognition; right) in Experiment 1. Results are shown separately as a function of age (older adults [Old] or younger adults [Yg]), category size (one or six category exemplars presented at study), and labeling condition (label–label [LL], clear–clear [CC], label–clear [LC], and clear–label [CL], in which the first term designates the condition at study and the second term designates condition at test). The results are shown without correction, and false alarms to novel items (baseline rates of false alarms when zero category exemplars were presented at study) are also shown. Error bars represent standard errors of the means.

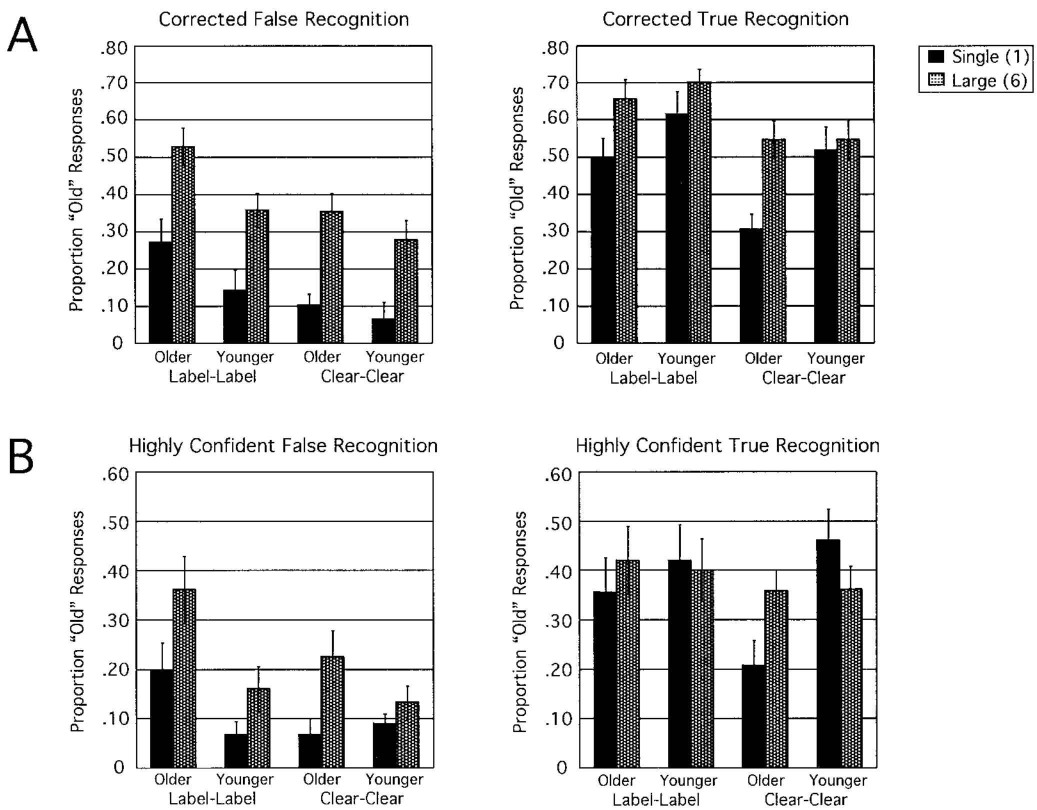

Figure 4. A: Proportions of "old" responses to nonstudied items (false recognition; left) and studied items (true recognition; right) after correction for baseline differences in false alarms to novel category lures, in Experiment 1. Outcomes are shown as a function of age and category size, separately for the label–label and clear–clear conditions. B: Mean proportions of highly confident "old" responses for the same conditions shown in A. Results are shown after subtracting highly confident false alarms to novel category items. Error bars represent standard errors of the means.

MSE = 0.05, $p = .008$, and a main effect of category size, with higher levels of false recognition for the large than for the single-item categories, $F(1, 68) = 68.67$, $MSE = 0.03$, $p < .0001$.

After correcting for novel false alarms, false recognition responses were still more common in older adults than in younger adults, $F(1, 68) = 6.51$, $MSE = 0.06$, $p = .01$, and for large categories than for single categories, $F(1, 68) = 68.67$, $MSE = 0.03$, $p < .0001$. However, there was now also a main effect of condition, reflecting higher novel-corrected false recognition in the LL than in the CC condition, $F(1, 68) = 9.51$, $MSE = 0.06$, $p = .003$, and no longer an Age × Condition interaction ($F = 1.31$).

To specifically test the prediction of the impaired perceptual encoding account, we contrasted false recognition of older and younger adults in the CC condition alone. Contrary to the expectations of the impaired perceptual encoding approach but consistent with the semantic categorization account, this analysis showed no effect of age on the raw false recognition scores ($F < 1$; *M*s = .37 and .35, for older adults and younger adults, respectively, for single and large categories combined) or for the novel-corrected false recognition measure ($F = 1.20$; *M*s = .23 and .17, for older adults and younger adults, respectively). Confining consideration to the large category lures alone likewise showed no effects (for the raw measure, $F < 1$; for the corrected measure, $F = 1.21$).

Veridical recognition. A 2 (age) × 2 (category size) × 2 (condition) ANOVA on the raw veridical recognition scores showed a main effect of age, with older adults demonstrating lower true recognition than younger adults, $F(1, 68) = 5.67$, $MSE = 0.06$, $p = .02$, and of category size, with higher true recognition for the large than for the single-item categories, $F(1, 68) = 22.11$, $MSE = 0.03$, $p < .0001$; both were modified by an Age × Category Size interaction, $F(1, 68) = 6.92$, $MSE = 0.03$, $p = .01$. This interaction reflected a greater single-category recognition disadvantage in the older group than in the younger group (a single category disadvantage of 20% vs. 6%, respectively). Each of these three effects was also found after correcting for novel false alarms. However, in addition, there was an effect of condition, $F(1, 68) = 9.69$, $MSE = 0.07$, $p = .003$, reflecting higher true recognition in the LL than in the CC condition.

High confidence responses. Focusing on the false recognition responses given with high confidence (that is, those accompanied by ratings of 4 or 5 on the 5-point confidence scale), for the two extreme conditions (LL and CC), there was a numerically larger difference between these conditions for older adults. This pattern was found both for the raw high-confidence false recognition responses (older adults, CC = .21, LL = .30, difference of 9%; younger adults, CC = .13, LL = .12, difference of 1%) and for novel-corrected high-confidence false recognition responses (older adults, CC = .15, LL = .28, difference of 13%; younger adults, CC = .11, LL = .11, difference of 0%). As expected within the semantic categorization account, for the novel-corrected outcomes, there was a trend toward an Age × Labeling Condition interaction, $F(1, 68) = 2.89$, $MSE = 0.05$, $p = .09$, reflecting an increase in highly confident false recognition responses by the older adults but not by the younger adults in the LL condition compared with the CC condition.

Turning to high-confidence true recognition, there was only one significant effect: an Age × Category Size interaction found both for the raw data and for the novel-corrected outcomes, $F(1, 68) = 8.31$, $MSE = 0.03$, $p = .005$. This interaction reflected a 11% true recognition difference in favor of many-exemplar items for older adults but a 6% difference in favor of single-exemplar items for the younger adults.

Discussion

The aim of this experiment was to test the semantic categorization versus the impaired perceptual encoding accounts of the increased vulnerability to false recognition responding in older compared with younger adults while holding the type of perceptual information constant. To this end, we developed a set of ambiguous perceptual stimuli for which the concurrent presence of semantic information could be manipulated. The strongest test of the two accounts was provided by a comparison of the false recognition rates in the LL condition (which involved presentation of the disambiguating label at both encoding and retrieval) with those in the CC condition (which involved no labels at all). Within the semantic categorization account, compared with younger individuals, older adults should have shown elevated false recognition in the LL condition but not in the CC condition. In contrast, the impaired perceptual encoding hypothesis predicted elevated false recognition for older adults in both conditions. The outcomes for the raw false recognition measure were consistent with the semantic categorization account: There was a reliable Age × Labeling Condition interaction for these two extreme conditions, such that older adults showed particularly elevated false recognition in the LL condition. In addition, contrary to what was expected within the impaired perceptual encoding account, a direct comparison of the levels of false recognition for the two age groups in the CC condition, in which no labels were presented, showed no age differences, either overall (combining across category size) or for large-category lures alone, and neither for raw nor novel-corrected measures.

However, because the labeling manipulation also exerted an effect on novel false alarms, support for the semantic categorization account was not found after correction for novel false alarms. (Note, however, that a clear trend did still emerge in the high-confidence responses, even after correction, with older adults demonstrating an increase of 13% in novel-corrected high-confidence responses in the LL condition compared with the CC condition, but with younger adults showing no difference.) With hindsight, this systematic effect of the experimental manipulation on baseline false alarms most likely reflected differences in the number of cues to novelty that could be exploited at the time of test for the different conditions (see also Koutstaal & Schacter, 2001; Schacter, Israel, & Racine, 1999). Whereas in the conditions in which no labels had been presented at study, only the visual figure itself could be used as the basis for a correct rejection of a novel lure, in conditions involving labels at study, new items could be correctly rejected either because the figure was judged to be novel or because the label (which was newly presented or inferred at test) was judged to be novel.

One further difficulty with interpretations drawn from the labels manipulation is that the labels might also have exerted an effect on the perceptual similarity of the items. It might be argued that the presence of labels, although encouraging semantic and conceptual processing, may also have made the labeled items more similar to one another, not only conceptually (through designation of the semantic category) but also perceptually (through introducing the shared similarity of the same visual letter string). Moreover, this might be especially true for the large category condition. Because of these two interpretative problems, we conducted a further experiment involving both abstract and concrete items within the same experiment and within participants.

Experiment 2

The approach in Experiment 2 was simple. First, visual arts students created colored line drawings of two large sets of new stimuli: one set depicting common everyday objects and the other depicting novel or abstract objects without preexisting semantic representations. Then, in the experiment, older and younger participants were presented with both abstract and concrete objects in one (randomly intermixed) study phase, with some categories represented by many exemplars and others by only a single exemplar per category. This study phase was followed by an old–new recognition test (with confidence ratings) in which, again, abstract and concrete items were randomly intermixed.

According to the semantic categorization account, whereas older individuals when tested for the common objects should show elevated false recognition relative to younger adults, this pattern should not be found—for these same individuals and relative to the same comparison group of younger adults—for abstract objects. In addition, given the patterns observed in earlier experiments, it might be expected that this differential pattern manifested as an Age × Stimulus Type interaction in false recognition responses should be particularly clear for a comparison restricted to those lure items that were endorsed with a high level of confidence. The impaired perceptual encoding account predicts that false recognition responses in the older group should be elevated above the younger group, regardless of the nature of the stimuli (i.e., abstract or concrete).

Method

Participants. A total of 18 older and 18 younger adults took part in the experiment. Older and younger individuals were recruited, screened, and reimbursed in the same manner as in Experiment 1. The average age of older participants was 68.1 years ($SD = 4.10$, range = 60–75 years); they reported an average of 16.5 years ($SD = 2.77$) of formal education. The average age of the younger participants was 19.2 years ($SD = 1.44$, range = 17–22 years); they had, on average, 13.1 years ($SD = 1.28$) of formal education. Older adults had significantly more years of education than did younger adults, $F(1, 34) = 22.20$, $MSE = 4.66$, $p < .0001$.

Design. The experimental design included a between-subjects variable of age (older vs. younger) and two within-subjects variables: stimulus type (abstract or concrete objects) and category size. For studied items, category size had three levels: single (targets for which one categorically related item was presented during study), large (targets for which nine categorically related items were presented during study), and unrelated (noncategory items, targets that were unrelated to other items at study or test). For nonstudied items, category size had four levels: single (related lure items for which one categorically related item had been presented at study), large (related lure items for which nine categorically related items had been presented at study), unrelated (unrelated lures that were unrelated to any items at study or test), and novel (lure items for which no categorically related items had been presented during study but from categories that across participants were systematically counterbalanced across study and test status). The novel category items provide an index of the baseline level of false alarms to categorized items.

Stimulus materials. The stimuli were colored line drawings of abstract and common objects, manually created and colored by two visual arts students and then scanned and converted to appropriate format for computer screen display during testing (see Figure 5). The abstract objects were similar to the drawings of common objects in aspects such as level of complexity and three-dimensionality. Selection of the stimuli involved an iterative procedure of stimulus creation, collection of informal normative ratings, and modification of the stimuli. The stimuli were created to vary in their degree of similarity to a prototype for each category, but (in the interests of brevity) data relating to the perceptual distance manipulation are not reported here.

As in the earlier experiments, to avoid confounding the size of the category at study with the size of the category at test, we tested only a subset of the studied items, termed *critical items,* during recognition, thus holding category size at test constant while manipulating category size during encoding. Specifically, with the exception of the single-item categories, which were tested once with the target and twice with related lures (once with a lure from a middle distance and once with the prototype), each category was tested with three studied items (targets at each of the near, middle, and far perceptual distances) and four related lures (lures at each of the near, middle, and far perceptual distances, plus the prototype). To accomplish this, we selected six items from each category to serve as critical items. These were assigned to two equal subsets of three items each (Critical Subset A and Critical Subset B, each consisting of a near, middle, and far distance item). Across participants, these subsets were rotated across study–test status. Prototypes were presented only as lures and were presented for novel, single, and large categories.

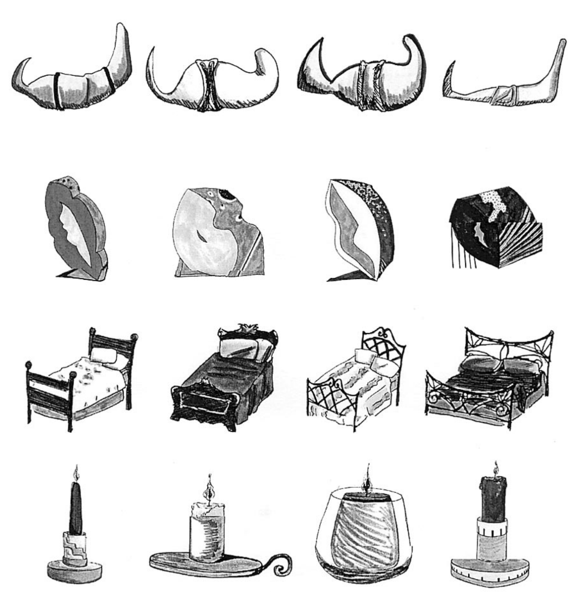

Figure 5. Examples of the stimuli used in Experiment 2. The top two rows show items from two abstract categories, and the lower two rows show items from two concrete categories. Note that although, for illustrative purposes here, the stimuli are shown in black and white, the stimuli as shown to participants were presented in color, with color being an important attribute that was varied both within and across categories.

Procedure. The overall procedure involved three phases, including a study phase in which participants were exposed to the stimuli under an incidental encoding task, a brief retention interval, and the test phase. All participants were tested individually in a single experimental session of approximately 75–90 min. The stimuli at both study and test were presented in the center of a color computer monitor, with prompts for responding to the encoding task or the recognition task displayed beneath.

In the study phase, each item was presented for 3 s, and participants were asked to rate how much they liked the visual appearance of the stimulus (1 = *do not like*; 5 = *very much like*). Each participant was exposed to a total of 190 items at study, including 9 items from each of 8 large abstract categories (72 items), 9 items from each of 8 large concrete categories (72 items), 1 item from each of 8 single abstract categories (8 items), 1 item from each of 8 single concrete categories (8 items), 12 unrelated abstract and 12 unrelated concrete items, plus 3 primacy and 3 recency buffers. The items during study were presented in a pseudorandom order such that items from the different category sizes and stimulus types were distributed equally throughout the list. The study phase was followed by a 10-min retention interval, during which participants performed unrelated filler tasks.

In the test phase, participants were shown a subset of the items that had been presented in the study phase, together with new items (both related lures and novel items). Participants were asked to indicate, using designated keys of the computer keyboard, whether each item was old (i.e., had been presented during the prior liking-rating task) or new (i.e., had not previously been presented in the experiment) and then to indicate their degree of confidence in their recognition judgments (1 = *just guessing*; 5 = *very sure*). Each test list consisted of a total of 272 items: 24 large category abstract targets (3 items × 8 categories, including near, middle, and far targets), 32 large category abstract lures (4 items × 8 categories, including near, middle, far, and prototype lures), 24 large category concrete targets (3 items × 8 categories), 32 large category concrete lures (4 items × 8 categories), 8 single category abstract targets (1 item × 8 categories), 16 single category abstract lures (2 items [including a middle distance lure and the prototype lure] × 8 categories), 8 single category concrete targets (1 item × 8 categories), 16 single category concrete lures (2 items × 8 categories), 32 novel category abstract lures and 32 novel category concrete lures (4 items × 8 categories each), 12 unrelated abstract targets, 12 unrelated abstract lures, 12 unrelated concrete targets, and 12 unrelated concrete lures. New and old test items were presented in a pseudorandom order such that items of the different stimulus types and category sizes were evenly distributed throughout the test.

Results

False recognition. Given the interpretive difficulties introduced by the differing patterns of baseline false alarms in Experiment 1, we first considered levels of false alarms to the novel category items. A 2 (age) × 2 (stimulus type) mixed-factor ANOVA revealed no overall age difference in the level of false alarms to novel items (M = .13 for both older adults and younger adults; $F < 1$). Averaging across the two age groups, there was an effect of stimulus type, with baseline false alarms for the abstract items (.17) exceeding those for the concrete items (.09), $F(1, 34) = 10.88$, $MSE = 0.01$, $p = .002$. However, this effect was modulated by age, such that whereas older adults showed identical levels of novel false alarms to the abstract and concrete items (.13 for both item types), younger adults showed elevated false alarms to the novel category abstract items (.20) compared with the novel category concrete items (.06), $F(1, 34) = 9.85$, $MSE = 0.01$, $p = .004$. Note, however, that this differing pattern in baseline false alarms will, if anything, lead to novel-corrected outcomes that tend to work against the pattern expected within the semantic categorization account. That account proposes that there will be no age difference in the level of false recognition for the abstract items. Subtracting a higher baseline from the abstract items for the younger adults should thus make any elevation in false recognition for the older adults that is found for abstract items even more pronounced. It will increase the difference between older and younger adults by depressing the rate of false alarms in the younger adults, thereby working contrary to the expected pattern of no or minimal age differences for these items. The novel-corrected recognition scores are presented in Figure 6.

A 2 (age) × 2 (category size) × 2 (stimulus type) ANOVA performed on the novel-corrected false recognition scores revealed a main effect of age, $F(1, 34) = 4.24$, $MSE = 0.02$, $p = .047$, reflecting a modest overall elevation in false recognition responses by older adults (.25) compared with younger adults (.19). As expected, there was also an effect of category size, reflecting much higher rates of false recognition for the many-exemplar category lures (.38) than for the single-item category lures (.06), $F(1, 34) = 254.63$, $MSE = 0.02$, $p < .0001$. This effect of category size was also moderated by age, such that older adults showed a larger increment in false recognition for the many-exemplar lures relative to the single-exemplar lures (increment of 37%) than did younger adults (increment of 27%), $F(1, 34) = 5.89$, $MSE = 0.02$, $p = .02$, for the Age × Category Size interaction. In addition, there was an overall effect of stimulus type, with the level of false recognition for concrete items (.25) exceeding that for abstract items (.19), $F(1, 34) = 8.72$, $MSE = 0.02$, $p = .006$. A crucial point, however, is that this difference was modified by age, with only older adults showing an increment for concrete items (14%) and younger adults showing similar levels of false recognition regardless of stimulus type (no increment), $F(1, 34) = 8.95$, $MSE = 0.02$, $p = .005$, for the Age × Stimulus Type interaction.[1]

Contrary to the outcomes expected within the impaired perceptual encoding account, analyses on the novel-corrected scores restricted to the abstract items alone showed no age effects, considering either all of the categorized items or the many-exemplar items alone (Fs < 1). For the raw false recognition responses, combining across category size, there was a main effect of age, but likewise contrary to the impaired perceptual encoding account, this effect arose from higher false recognition of the abstract items by younger adults (.40) than by older adults (.31), $F(1, 34) = 4.40$, $MSE = 0.03$, $p = .04$. This effect was modified by an Age × Category Size interaction, $F(1, 34) = 7.01$, $MSE = 0.01$, $p = .01$, reflecting an elevation in raw false recognition in the younger group only for the single items (Ms = .25 and .11, for the younger adults and the older adults, respectively) but not for the many-exemplar items (Ms = .54 and .51, for the younger adults and the older adults, respectively; for the age comparison restricted to the abstract many-exemplar items alone, $F < 1$).

Veridical recognition. A 2 (age) × 2 (category size) × 2 (stimulus type) ANOVA performed on the novel-corrected true recognition scores showed no overall age difference in veridical recognition (Ms = .55 and .61, for the older individuals and the younger individuals, respectively; $F < 1.3$). There was an overall effect of category size, with recognition of many-exemplar targets (.64) exceeding that for one-of-a-kind items (.52), $F(1, 34) = 18.08$, $MSE = 0.03$, $p = .0002$. This effect of category size, although not significantly moderated by age (for the Age × Category Size interaction, $F = 2.24$, $p = .14$), was numerically more apparent for older adults (difference of 17%) than for younger adults (difference of 8%). There was also a robust effect of stimulus type, with recognition of the concrete items (.68) well exceeding that of the abstract items (.47), $F(1, 34) = 53.62$, $MSE = 0.03$, $p < .0001$. In both older and younger adults, this effect of stimulus type was more pronounced for the one-of-a-kind items (concrete advantage of 27%) than for the many-exemplar items (concrete advantage of 15%), $F(1, 34) = 8.48$, $MSE = 0.01$, $p = .006$, for the Stimulus Type × Category Size interaction. (For the Age × Stimulus Type × Category Size interaction, $F < 1$.)

High-confidence responses. Consideration of only those false recognition responses accompanied by high levels of confidence showed several effects, including a main effect of age, with high-confidence false recognition responses more frequent among older adults (.18) than younger adults (.10), $F(1, 34) = 11.37$, $MSE = 0.02$, $p = .002$. There was also a main effect of category size, $F(1, 34) = 73.68$, $MSE = 0.02$, $p < .0001$, and an Age × Category Size interaction, $F(1, 34) = 6.63$, $MSE = 0.02$, $p = .01$, reflecting a nearly twofold greater increase in highly confident false recognition as a function of category size in older adults (increment of 24%) than in younger adults (increment of 13%). There was also a main effect of stimulus type, with high-confidence false recognition of concrete items (18%) nearly twice that for abstract (10%) items, $F(1, 34) = 20.00$, $MSE = 0.01$, $p < .0001$. A crucial point is that there was again an Age × Stimulus Type interaction. The increase in confident false recognition for concrete relative to abstract items was five times greater in older adults (concrete increment of 15%) than in younger adults (concrete increment of 3%), $F(1, 34) = 9.92$, $MSE = 0.01$, $p = .003$. This pattern was also very clear for the large category items alone (concrete increments of 17% for older adults versus 0% for younger adults), $F(1, 34) = 8.92$, $MSE = 0.02$, $p = .005$, for the Age × Stimulus Type interaction.

Consideration of highly confident veridical recognition showed that high-confidence recognition for concrete targets (.59) well exceeded that for abstract target items (.32), $F(1, 34) = 88.25$, $MSE = 0.03$, $p < .0001$. There was again a tendency for older adults to show a more pronounced effect of category size (one-of-a-kind deficit of 6%) than younger adults (one-of-a-kind advantage of 3%), but this interaction was not significant ($F = 2.74$, $p = .11$).

Finally, consideration of highly confident false recognition for the abstract items alone again provided no support for the impaired perceptual encoding account. There was no age difference for novel-corrected confident false recognition of the abstract items, either combining across category size ($F < 1.1$; Ms = .11 and .08,

[1] An analysis performed on the raw false recognition responses likewise showed the critical Age × Stimulus Type interaction, but now with a crossover pattern. Older adults showed an increment of 13% for concrete compared with abstract items, and younger adults showed an equivalent increment, but of the opposite pattern, with false recognition of abstract items exceeding that for concrete items by 14%, $F(1, 34) = 34.23$, $MSE = 0.02$, $p < .0001$.

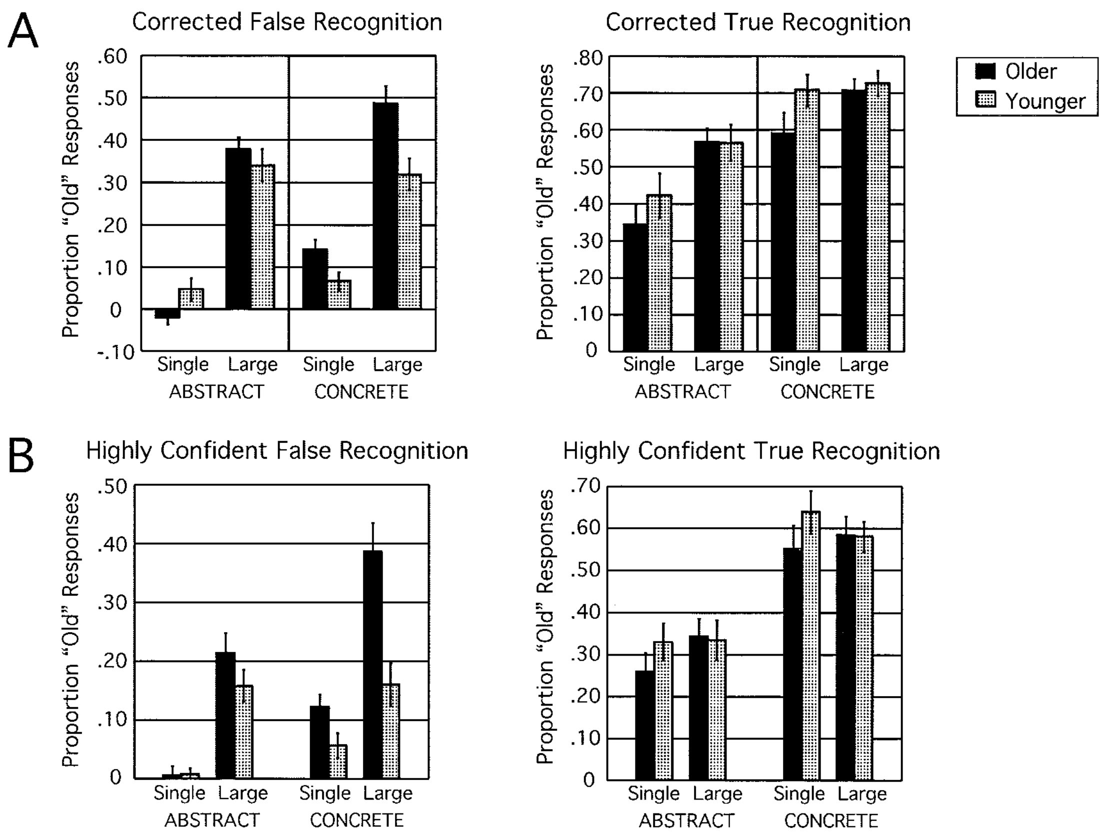

Figure 6. A: Mean proportions of "old" responses to nonstudied items (false recognition; left) and to studied items (true recognition; right) after subtracting false alarms to novel category items, for Experiment 2. Results are shown separately as a function of age (older adults or younger adults), category size (one or nine category exemplars presented at study, termed *Single* and *Large,* respectively), and stimulus type (abstract or concrete objects). B: Mean proportions of highly confident "old" responses for the same conditions shown in A. Results are shown after subtracting highly confident false alarms to novel category items. Error bars represent standard errors of the means.

for the older adults and the younger adults, respectively) or for the many-exemplar categories alone ($F = 1.62$; Ms = .21 and .16, for the older adults and the younger adults, respectively).

General Discussion

The level of false recognition in older and younger adults was examined in two experiments. Each experiment involved stimulus items that did versus did not possess preexisting conceptual or semantic representations. The results of Experiment 1, which involved the use of ambiguous objects accompanied or not accompanied by disambiguating conceptual labels at study and/or at test, provided support for the semantic categorization account. Older adults showed elevated false recognition in the condition involving labels at both study and test but not in the condition without labels (as shown by a significant Age × Condition interaction in raw false recognition and a trend toward an interaction in high-confidence false recognition after correction for false alarms). However, the outcomes were not entirely clear-cut, because the experimental manipulation of the labels also exerted systematic effects on the rates at which novel items were incorrectly endorsed as having been studied. These systematic effects in the baseline levels of false alarms eliminated the expected Age × Stimulus Type interaction in the overall recognition measure involving correction for these differences. In Experiment 2, which was conducted with a new set of abstract and common objects and designed to allow a within-subject test of the semantic categorization account, we addressed these concerns. The key outcome of Experiment 2 was the demonstration of especially heightened false recognition in older adults for concrete but not for abstract objects. This pattern, expected by the semantic categorization account, was found for all of the measures used (as shown by Age × Stimulus Type interactions in raw false recognition, in novel-corrected false recognition, and in highly confident false recognition).

In notable contrast, there was little support for the impaired perceptual encoding account of the age-related elevation in false positive responding. Within this account, older adults may less efficiently encode and later use the comparatively rich and detailed perceptual information provided by visual stimuli, whether concrete or abstract. In Experiment 1, there was no age-related increase in false recognition in the CC condition (the condition in which the stimuli most nearly approximated those of novel abstract objects), either for raw false recognition or for the novel-corrected outcomes. Likewise, in Experiment 2, older adults did not show elevated false recognition for the abstract items, either for the novel-corrected outcomes or for highly confident responses; indeed, for raw false recognition responses to single items, younger adults actually showed greater false recognition than did older adults.

The conclusion that older adults may not always show impairments in memory for perceptual information coheres nicely with the results of a recent study (Koutstaal, 2003). That study examined the effects of across-exemplar changes in perceptual detail on three memory tasks, for which such changes might have intentional versus unintentional effects on performance. Older and younger adults were equally affected by across-exemplar changes on two tasks in which perceptual information might affect memory retrieval in indirect or nonintentional ways (in a repetition priming task and a meaning-based recognition task, for which episodic recognition decisions were to be made not on the basis of identity but on the basis of similarity in concept or meaning). This was so even though—for the same participants and precisely the same encoding conditions—the older adults showed, as expected, higher false recognition than did the younger adults on a standard episodic recognition test. This pattern suggests that perceptual details were indeed encoded by the older adults but were less effectively used in the recognition context requiring their deliberate, controlled use.

Taken together, the experiments reported here demonstrate that there is a comparative (although not absolute) age-related difference in susceptibility to gist-based errors as a function of the type of stimulus information that is present, with older adults more strongly adversely affected when semantic or conceptual information is present. These experiments also have clarified some of the parameters involved in this exacerbation. For example, the exacerbation of false positive responses in the presence of semantic information is not simply attributable to differences in very low-grade forms of familiarity, because age differences emerged at least as strongly, or more strongly, in positive recognition responses that were accompanied with high confidence. In addition, the exacerbation may require the presence of semantic information both during initial processing and during attempted retrieval (see Figure 3), and it is generalizable across different types of materials. (These experiments and the preliminary studies each used different stimuli but consistently pointed toward the same conclusion.)

Finally, we might underscore the comparatively greater clarity of the outcomes observed in these and the preliminary experiments when considering recognition responses endorsed with high confidence rather than (or in addition to) all responses. On the one hand, particularly when focusing on possible age-related differences in memory performance, the relatively reduced level of false recognition in the younger group for highly confident responses points to the incremental validity achieved by probing people's degree of confidence in a memory judgment. It points to the successful application, in some instances, of a second filtering or monitoring of people's decisions (see also Koriat & Goldsmith, 1996; for discussion in relation to recent neuroimaging findings, see Henson et al., 2000). On the other hand, when focusing less on the age-related differences and simply noting the frequency with which similar-seeming lure items were mistakenly endorsed with high levels of confidence, regardless of age, our conclusions may be somewhat less sanguine. We can be confident but wrong, if our justification for confidence is itself derived from information that does not allow differentiation (see also Chandler, 1994; Migueles & García-Bajos, 1999). This misplaced mnemonic confidence is more likely to occur with increasing age and more likely to occur when people have previously encountered many similar items, particularly when those items possess preexisting semantic representations. Yet, even in the most favorable conditions, occasional instances of misplaced mnemonic confidence are seldom completely averted.

References

Anderson, N. D., & Craik, F. I. M. (2000). Memory in the aging brain. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 411–425). Oxford, England: Oxford University Press.

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology, 16,* 361–384.

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395–409). Oxford, England: Oxford University Press.

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

Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to "remember" than things that were. *Psychological Science, 9,* 484–489.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition, 22,* 273–280.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments, and Computers, 25,* 257–271.

Craik, F. I. M. (1982). Selective changes in encoding as a function of reduced processing capacity. In F. Klix, J. Hoffman, & E. van der Meer (Eds.), *Cognitive research in psychology* (pp. 152–161). Amsterdam: North-Holland.

Henson, R. N. A., Rugg, M. D., Shallice, T., & Dolan, R. J. (2000). Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. *Journal of Cognitive Neuroscience, 12,* 913–923.

Hess, T. M. (1982). Visual abstraction processes in young and old adults. *Developmental Psychology, 18,* 473–484.

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

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.

Jackson, E. M. (1998). *Age-related differences in memory: False recognition and prototype abstraction in older and younger adults.* Unpublished Honors Bachelor of Arts thesis, Harvard University, Cambridge, MA.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517.

Koutstaal, W. (2000). [Results of follow-up experiment to Jackson (1998)]. Unpublished raw data.

Koutstaal, W. (2003). Older adults encode—but do not always use—

perceptual details: Intentional versus unintentional effects of detail on memory judgments. *Psychological Science, 14,* 189–193.

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

Koutstaal, W., & Schacter, D. L. (2001). Memory distortion and aging. In M. Naveh-Benjamin, M. Moscovitch, & H. L. Roediger III (Eds.), *Perspectives on human memory and cognitive aging: Essays in honour of Fergus Craik* (pp. 362–383). Philadelphia, PA: Psychology Press.

Koutstaal, W., Schacter, D. L., & Brenner, C. (2001). Dual task demands and gist-based false recognition of pictures in younger and older adults. *Journal of Memory and Language, 44,* 399–426.

Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging, 14,* 220–237.

Koutstaal, W., Schacter, D. L., Verfaellie, M., Brenner, C., & Jackson, E. M. (1999). Perceptually based false recognition of novel objects in amnesia: Effects of category size and similarity to category prototypes. *Cognitive Neuropsychology, 16,* 317–341.

LaVoie, D. J., & Faulkner, K. (2000). Age differences in false recognition using a forced choice paradigm. *Experimental Aging Research, 26,* 367–381.

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.

Macromedia, Inc. (1996). Aldus FreeHand 7.0 [Computer software]. San Francisco, CA: Author.

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

Migueles, M., & García-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology, 13,* 257–268.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25,* 838–848.

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

Park, D. C., Puglisi, J. T., & Sovacool, M. (1983). Memory for pictures, words, and spatial location in older adults: Evidence for pictorial superiority. *Journal of Gerontology, 38,* 582–588.

Park, D. C., Puglisi, J. T., & Sovacool, M. (1984). Picture memory in older adults: Effects of contextual detail at encoding and retrieval. *Journal of Gerontology, 39,* 213–215.

Reddy, C. (2000). *The role of conceptual labeling in age-related false recognition.* Unpublished Honors Bachelor of Arts thesis, Harvard University, Cambridge, MA.

Robinson, K. J., & Roediger, H. L. III. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8,* 231–237.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40,* 1–24.

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

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49,* 289–318.

Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition, 27,* 538–552.

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

Solso, R. L., & Raynis, S. A. (1979). Prototype formation from imaged, kinesthetically, and visually presented geometric figures. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 701–712.

Trahan, D. E., Larrabee, G. J., & Levin, H. S. (1986). Age-related differences in recognition memory for pictures. *Experimental Aging Research, 12,* 147–150.

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

Winograd, E., Smith, A. D., & Simon, E. W. (1982). Aging and the picture superiority effect in recall. *Journal of Gerontology, 37,* 70–75.

Received October 26, 2001
Revision received December 29, 2002
Accepted January 15, 2003 ■