A ge-R elated D ifferences in the U se of Contextual I nformation in R ecognition M emory: A G lobal M atching A pproach

U te J . Bayen, Matthew P. Phelps, and Julia Spaniol

1Department of Psychology, The University of North Carolina at Chapel Hill.
2Department of Psychology, Malone College, Canton, Ohio.

Age differences in the processing of contextual information were investigated using the Item, associated Context, and Ensemble (ICE) model (K. Murnane, M. P. Phelps, & K. Malmberg, 1999), a general global matching model of recognition memory. In two experiments, young and older adults studied words in environmental contexts and were tested in both the same and different contexts. Patterns of context effects for hit rate, false alarm rate, and $d^\prime$ suggest that older adults process associated context, but have difficulties integrating items and context into an ensemble. Thus, older adults appear to have a specific, rather than a general, deficit in processing contextual information. A deficiency in ensemble processing may be responsible for the prevalent finding that older adults show poorer recognition memory performance than young adults.

ONE of the best-established results of research in the area of human memory and aging is that older adults' performance on episodic-memory tasks is generally lower than that of young adults (for reviews see Craik & Jennings, 1992; Kausler, 1994; A. D. Smith, 1996). Episodic memory is a term coined by Tulving (1972) that denotes memory for specific events. An episode is defined by the context (e.g., time, environment, mood state) in which information is encoded. For example, if you try to remember whether you had muffins for breakfast this morning, it is not sufficient for you to know what muffins are or that you usually eat muffins, rather you have to remember if muffins occurred in the particular context of your breakfast this morning. Because successful performance on episodic-memory tasks requires the encoding and retrieval of the context in which information was acquired, gerontologists have suggested that older adults’ lower performance on such tasks might be attributable to age-related differences in the processing of context information. This hypothesis was first proposed in the early 1980s (Burke & Light, 1981; Craik & Byrd, 1982; Rabinowitz, Craik, & Ackerman, 1982) and has since inspired a number of researchers to investigate the role of context processing in age differences on episodic-memory tasks.

In this article we examine the effects of changes in environmental context between learning and test on recognition memory in young and older adults. Although age differences in recall are generally larger than age differences in recognition (Craik & McDowd, 1987), statistically significant age differences in recognition were found in most studies (for a review, see Kausler, 1994). We propose that older adults have a specific, rather than a general, deficit in processing contextual information in recognition tasks. More specifically, we argue that older adults process contextual information in a nonintegrative manner as efficiently as younger adults, but have difficulties integrating context information with to-be-remembered items. A specific deficit in integrating to-be-remembered items with their surrounding context may provide an explanation for the prevalent finding that older adults show poorer recognition memory performance than young adults.

One line of research that investigates the issue of context processing in older adults is research on aging and source memory, which involves direct questions about the source of information or the context in which information was acquired (e.g., Bayen & Murnane, 1996). In a source-memory paradigm, participants study information in different contexts and are later asked to remember the context in which particular pieces of information were learned. According to the results of this research, under most circumstances older adults show lower performance on such tasks than young adults (see a meta-analysis by Spencer & Raz, 1995), indicating that older adults’ memory for context is impaired.

Another line of research investigates the facilitative effects of context memory on memory for the content of episodes. If, in comparison to young adults, older adults are less adept at encoding and retrieving contextual information, then they are expected to show less encoding specificity. One form of encoding specificity (Tulving & Thompson, 1973), a context effect, refers to the effect that memory performance is better the more similar contextual conditions are between encoding and retrieval. If older adults have difficulties encoding or retrieving context, their memory for items learned in a particular context should benefit less than that of young adults from matching contextual conditions at encoding and retrieval.

Several studies have compared context effects in young and older adults. Using recognition tasks, Naveh-Benjamin and Craik (1995) found that older adults benefit at least as much as their younger counterparts when environmental
context at encoding matches that at test. An absence of age differences in environmental-context effects on recognition was also reported by Vakil, Melamed, and Even (1996). Similarly, Schramke and Bauer (1997) found age-independent context effects of physiological states on free recall, and Light, LaVoie, Valencia-Laver, Albertson Owens, and Mead (1992) reported increased repetition priming in both young and older adults with matching versus nonmatching acquisition and test modalities. Park and collaborators also did not find age differences in the effects of encoding specificity using picture recognition (Park, Puglisi, Smith, & Dudley, 1987; Park, Puglisi, & Sovacool, 1984) and word recall (Puglisi, Park, Smith, & Dudley, 1988).

However, Park, Smith, Morrell, Puglisi, and Dudley (1990) as well as A. D. Smith, Park, Earles, Shaw, and Whiting (1998) asserted that older adults do show an impairment in context processing when a task requires self-initiated integration of item and context. They concluded this from their findings of larger age differences in cued-recall when items and their contextual cues were unrelated to each other than when they were related to each other. In a study by Earles, A. D. Smith, and Park (1996), older adults benefited from instructions to integrate item and context information, but not as much as a comparison group of young adults.

It thus appears that older adults do encode and retrieve context information, but that they have problems with the integration of item and context information. In our research, we have tested this hypothesis using a qualitative, formal theory of recognition memory that allows us to disentangle possible age differences in different forms of context processing.

Although formal models are available that give sophisticated accounts of the role of context in episodic-memory tasks such as recognition and recall tasks, none of these models have thus far been used to shed light on the issue of age differences in episodic memory. A model that is particularly suitable for cognitive-aging research is the Item, associated Context, and Ensemble (ICE) model developed by Murnane, Phelps, and Malmberg (1999), because it specifies two different kinds of context processing, namely, the processing of context separate from an item (associated context), and the formation of an integration of item and context information (an ensemble). Depending on which kind of context processing participants use, different patterns of context effects are predicted in recognition tasks.

We first describe the characteristics of the ICE model and summarize prior research on young adults that has validated this model. We then derive predictions regarding age differences in different forms of context processing and present two experiments to support our hypotheses.

The ICE Model of Recognition Memory

The ICE model (Murnane et al., 1999) is a general global matching model of recognition memory, in that it incorporates the common, general characteristics of the set of specific global matching models. Examples of global matching (or “global activation”) models of recognition are SAM (Gillund & Shiffrin, 1984), MINERVA2 (Hintzman, 1988), TODAM2 (Murdock, 1993), and the Matrix model (Humphreys, Bain, & Pike, 1989). For a review of global matching models see Clark and Gronlund (1996). In a recognition task, participants are presented with a list of items and are later asked which items from a test list appeared on the earlier study list. Global matching models assume that recognition of items in such a task is based on the activation of a large number of memory representations. These representations are activated in response to a retrieval cue.

For example, let us assume a research participant has studied a list of words, and then receives an old–new recognition test in which he or she has to indicate if a test item (e.g., “table”) had appeared on the study list or not. The participant forms a retrieval cue of test-item information and test context and matches this cue against the representations of all items and their contexts in memory. The degree to which a memory representation is activated by the retrieval cue depends on the similarity between the information in the cue and the information stored in memory. These component activations are combined (i.e., summed) to yield a global match, which forms the basis for the recognition judgment. The higher the global match, the greater the probability of an “old” response.

According to ICE, item information (I) plays a role in recognition as well as in the two theoretical forms of context information. These two forms of context are associated context (C), and ensemble (or integration) of item and context (E). These three components (I, C, and E) can each contribute to the global match. Consider the example of the word “rose” presented in the picture of a living room for a recognition test. In this example, any information encoded about “rose” that is not related to the living room context is item information (I), any information encoded about the picture of the living room that is not related to roses is associated-context information (C). Ensemble information (E) is a third and distinct type of information produced by integrating item and associated-context information. For example, you might integrate item and context into an ensemble by thinking of your significant other giving you a rose in the living room.

ICE makes predictions about context effects on different measures of recognition, including hit rate, false alarm rate, and the discrimination measure d’. Hit rate (HR) is the proportion of old items identified as old. False alarm rate (FAR) is the proportion of new items called old. Measures of discrimination (such as d’) are composite measures of recognition memory that index an individual’s ability to distinguish old from new items, independent of response biases.

In our paradigm, context effects are obtained in a list-learning experiment with an old–new recognition test in which items are either tested in the same context in which they were studied or in a different context that the participant did not see during study. For example, if the word “bear” was presented on a yellow background on the study list, it is either presented on the same yellow background at test or on a different background, such as a red background. Similarly, distractor items are either tested in a context that the participant experienced during the learning phase, or in a new, different context that was not presented during the learning phase of the experiment. A context effect occurs when HR, FAR, or d’ in same-context testing is larger than HR, FAR, or d’, respectively, in different-context testing.
The magnitude of a context effect is calculated by subtracting different-context scores from same-context scores.

The predicted patterns of context effects are different depending on which information is stored in memory and used in the retrieval cue, either $I$ and $C$ only, or $I$, $C$, and $E$. The formal derivations of the following predictions from the ICE model can be found in Murnane and collaborators (1999). ICE predicts that if only item information ($I$) and associated-context information ($C$) are used in the global match, then the global match for target items tested in the same context is higher than the global match for targets tested in a different context. Consequently, context effects are expected on HR. Also, the global match for distractors tested in a context that the participant has experienced during encoding (“same context”) is predicted to be higher than the global match for distractors presented in a new (“different”) context. Thus, context effects are also predicted on FAR. Because an associated-context match increases the global match to both targets and distractors, it is not likely to enhance discrimination as measured by $d'$. Thus, if only $I$ and $C$ are used, we do not predict a context effect on $d'$. (see Appendix, Note 1).

If, on the other hand, people use item information ($I$), associated-context information ($C$), and an ensemble ($E$) in the global match, then the ensemble is an additional source of match for targets tested in the same context. Consequently, the global match for targets tested in the same context is increased compared with the case in which no ensemble is used. The global match for distractors tested in a formerly presented context, however, is not increased compared with the no-ensemble case, because an ensemble cue formed from a distractor does not match any ensembles stored in memory. Because an ensemble match increases the global match to targets but not to distractors, it contributes to enhanced recognition memory as indicated by the discrimination measure $d'$. Thus, when $I$, $C$, and $E$ are used, context effects are predicted in HR, FAR, and $d'$.

In a series of seven experiments with young college students, Murnane and Phelps (1993, 1994, 1995) varied simple visual contexts (as defined by foreground color, background color, and location on a computer screen) of words at study and test and found context effects on both HR and FAR. In the majority of conditions, no context effects on $d'$ appeared. This research suggests that simple visual context is not conducive to the formation of an ensemble of item and context information. Thus, when simple visual contexts are used, only item information ($I$) and associated-context information ($C$) are encoded in memory and in the retrieval cue.

Murnane and collaborators (1999) found the pattern of results predicted by ICE when $I$, $C$, and $E$ contribute to the global match. They used a contextual manipulation that they referred to as rich visual context (i.e., pictures of scenes like those in Figure 1) and samples of young adult participants. Their results indicate that young adults encode rich visual context as both associated context and as context integrated with items into an ensemble. Unlike simple visual context, rich visual context is rich in meaningful content that can be integrated with meaningful items (i.e., words) into an ensemble.

In summary, we predict on the basis of the ICE theory and prior research (Murnane & Phelps, 1993, 1994, 1995; Murnane et al., 1999) that if only $I$ and $C$ are encoded in memory and used as retrieval cues, context effects will be observed in HR and FAR, but not in $d'$. If, however, $I$, $C$, and $E$ are used, context effects are predicted for HR, FAR, and $d'$.

In our research, we used the ICE model to shed light on possible age-related differences in context processing in recognition tasks. On the basis of prior research that has found evidence that older adults encode and retrieve context information in episodic memory (Light et al., 1992), we hypothesized that older adults process associated context ($C$) as effectively as do young adults. Prior research also indicates, however, that older adults have difficulties with the self-initiated integration of item and context information (Park et al., 1990; A. D. Smith et al., 1998). Within the framework of the ICE theory, this translates into an age-related impairment in the formation of an ensemble of item and context ($E$). We hypothesize that older adults have difficulties forming an ensemble from item and context information. Thus, in situations in which young adults use all three components of the ICE model (i.e., $I$, $C$, and $E$), older adults use only $I$ and $C$. Because it is the ensemble match that clearly contributes to enhanced recognition memory performance (i.e., discrimination between targets and distractors), an impairment in ensemble formation is a possible explanation for the prevalent findings of age-related differences in recognition memory.

Overview of the Experiments

To test our hypothesis that older adults use associated-context information in recognition tasks but have difficulties forming ensembles, we performed two experiments. In both of them, participants were presented with words in particular environmental contexts, which was followed by an old-new recognition test in which items were tested either in the same context in which they had been presented at learning or in a different context. Our goal was to study age differences under conditions that reveal the default mode of context processing, that is, how contextual information is
processed when conditions do not elicit complex strategies for remembering or for context integration. Therefore, we used an incidental learning procedure and an orienting task that did not explicitly require the formation of ensembles. An important rationale for investigating context processing under these sorts of conditions is that they mirror the conditions under which most episodic memories are formed in everyday life. Another reason for doing so is that it allows us to begin to answer the more basic question of how these populations tend, by default, to process contextual information before considering (in future research) the more complex issue of how they may be able to process context more effectively when they are instructed to do so.

In Experiment 1, we used simple visual contexts (combinations of foreground color, background color, and screen location of words), which as shown by Murnane and Phelps (1993, 1994, 1995) do not foster the formation of an ensemble of item and context information in young adults. In this task, observed context effects are solely based on item and associated-context information (I and C). If older adults process associated-context information, they should show a pattern of context effects similar to that of young adults. That is, both age groups should show context effects in both HR and FAR, but not in \( d' \). In Experiment 1, both age groups should show this same pattern of context effects.

For Experiment 2, we used a design similar to that of Experiment 1, with the difference that rich visual contexts (i.e., pictures of scenes) were used instead of simple visual contexts. Murnane and collaborators (1999) found that young adults encode associated context and integrate item and context into ensembles when rich visual context is manipulated between study and test. If, as our hypothesis predicts, older adults do not form and/or use an ensemble to the extent that young adults do, then young and older adults should show different patterns of context effects in this rich-context experiment. The data pattern for older adults in this experiment should be similar to the data pattern for young adults under conditions of simple visual contexts. That is, older adults should show context effects on HR and FAR, but not on \( d' \). On the other hand, young adults in Experiment 2 (rich visual contexts) should, in replication of results reported by Murnane and collaborators (1999), show context effects on HR, FAR, and \( d' \).

According to Murnane and Phelps (1993, 1994, 1995), an experimental design has to satisfy certain methodological conditions to make the above predictions of the ICE model applicable. First, the same-versus-different-context variable must be manipulated within participants and within the recognition-test list (as opposed to between participants or lists) to minimize the likelihood that different response criteria are adopted in response to different overall levels of activation in the two test-context conditions. Such differences in response criteria could mask context effects on HR and FAR (for a detailed discussion of these issues, see Murnane & Phelps, 1993). Second, the experiment should use what Murnane and Phelps (1994) refer to as an AB-X paradigm, as opposed to an AB-A paradigm. In an AB-X paradigm, the “different” test context is a context that was never experienced during study. In an AB-A paradigm, the “different” context was experienced during study, but is different from the context in which the test item was experienced during study. Murnane and Phelps’ (1994) formal derivations show that for an AB-X design, ICE predicts context effects on HR and FAR, whereas no or very small context effects are predicted if items are tested in an AB-A design. These different predictions have found empirical support (Murnane & Phelps, 1994).

We have designed two experiments in accordance with the methodological recommendations given by Murnane and Phelps (1993, 1994, 1995). That is, we have used AB-X designs in which the test-context variable is manipulated within participants and within the test list.

**Experiment 1**

In this experiment we used simple visual contexts. We predicted that the same pattern of context effects would emerge for both young and older adults. Both age groups should show context effects on HR and FAR; neither age group should show a context effect on \( d' \). This pattern of results is predicted if both age groups use associated context (C) in their recognition judgments, but not an ensemble of item and context (E).

**Method**

Participants.—A total of 104 adults participated in this experiment. Of these, 52 were young adult undergraduate students at The University of North Carolina at Chapel Hill (23 men, 29 women). They were between the ages of 18 and 24 years (mean age 19.7 years) and were recruited from introductory psychology courses. They received class credit as compensation for their participation in the experiment. Fifty-two of the participants were older adults (18 men, 34 women) between the ages of 59 and 83 years (mean age 71.0 years), resided in communities in central North Carolina, and were recruited through newspaper advertisements as well as flyers on community bulletin boards and in Senior Centers. They received payment for their participation in the study.

All participants were native speakers of English. Mean years of education were 13.3 (\( SD = 0.85 \)) for young adults and 15.4 (\( SD = 2.57 \)) for older adults, a significant difference, \( t(102) = 5.79, p < .0005 \). The mean performance in raw score units on the 30-point Gardner–Monge (Gardner & Monge, 1977) vocabulary test was 15.4 (\( SD = 3.67 \)) for young adults and 23.0 (\( SD = 4.56 \)) for older adults, also a significant difference, \( t(101) = 9.27, p < .0005 \). One young adult achieved only three points in the vocabulary test. This participant’s response pattern indicated that she had not taken this test seriously, and her vocabulary test score was therefore not included in the mean reported above. In a near-vision test, all participants could fluently read printed text in 7-point font size.

On a 4-point overall health rating scale with the options “poor,” “fair,” “good,” and “excellent,” 86% of the young adults and 98% of the older adults rated their health as good or excellent. None of the participants rated their health as poor. Persons who, in a phone interview, reported a history of heart attack, stroke, diabetes, emphysema, Parkinson’s disease, brain trauma, or alcoholism were excluded from participation in the study. Individuals who reported being
colorblind were also excluded. The data from two young adults were not logged because of technical failure. We excluded participants who did not give responses within the maximum response time of 5 s in four or more of the 64 recognition-test trials. Five participants were excluded for this reason: One older adult did not respond in 11 of the trials, 2 older adults did not respond in 5 of the trials, 1 young adult did not respond in 8 of the trials, and 1 young adult did not give any responses in the memory test. All excluded participants were replaced by additional participants.

Design.—The design was a 2 × 2 mixed factorial with age group (young vs. older) as a between-subjects variable, and test context (same as learning context vs. different-from-learning context) as a within-subjects variable.

Materials.—Stimulus items were 96 high-frequency nouns (Francis & Kucera, 1982). These items were presented on color monitors in 24-point-size lower-case letters. Two different combinations of item color, screen color, and screen location served as learning and test contexts for the items. In Context A, light-green items were presented in the upper left-hand corner of a magenta computer screen. In Context B, yellow items were presented in the lower right-hand corner of a blue screen.

Procedure.—Each experimental session included 1 or 2 older participants or between 1 and 4 young participants. After signing consent forms and taking a brief vision test, participants were directed to individual computer booths. They were informed that they would see word pairs on the computer screen, one pair at a time, and for each pair were supposed to rate on a 5-point scale how related they thought the two words were. Following the experiments by Murnane and collaborators (Murnane & Phelps, 1993, 1994, 1995; Murnane et al., 1999), we presented word pairs at study, instead of single words, to reduce interitem associations across study trials and cross-context encoding of items. For each trial, two words were randomly drawn from the pool of 96 nouns. Participants were not informed of the upcoming memory test. The purpose of the relatedness judgment task was to assure that participants focused their attention on and processed the items on the study list. A practice list of 8 word pairs was presented first. Participants were then presented with a study list of 32 word pairs. The two members of each word pair on the practice and study lists were presented side-by-side. Half of the participants of each age group studied the word pairs in Context A, and the other half studied them in Context B. Participants saw each word pair for 5 s, then the words disappeared and the sentence “Please enter your relatedness judgment now” appeared in the middle of the computer screen. This sentence served as a signal for participants to rate the relatedness of the word pair by pressing the number keys 1 through 5 on the computer keyboard (with 1 indicating least related, and 5 indicating most related). The next word pair appeared immediately after a response had been entered for the previous one or 5 s after the disappearance of the previous pair, whichever occurred first. The presentation order of the word pairs was randomized by participant.

After completion of the relatedness judgments, participants were informed of the upcoming recognition test. For this test, the computer keys “D” and “K” were labeled with stickers “YES” and “NO.” The assignment of computer keys to response options was counterbalanced across participants. Participants were instructed to use the digits of their right and left hands to hit the two response keys. They were furthermore instructed to hit the “YES” key if the test item had previously been presented on the study list and to hit the “NO” key if the test item had not appeared on the study list. Participants were asked to respond as fast and as accurately as possible to each test item. The recognition test was preceded by six practice trials after which participants had the opportunity to ask the experimenter questions. Again, this opportunity was rarely used. The test word disappeared immediately after the participant responded, and the next trial started after an intertrial interval of 150 ms during which a black, blank screen appeared. If a participant did not respond within 5 s after appearance of a test word, the word disappeared, and the next test trial started. Error feedback was not provided.

Thirty-two target items and an equal number of distractor items were presented on the recognition test. The order of test-item presentation was randomized by participant. From each of the 32 word pairs that had been presented at study, one item was randomly selected to serve as a target item in the recognition test, with the constraint that half of the target items had been the left member of a studied word pair, and the other half had been the right member of a pair. For each participant, items from the word pool that had not been presented on the study list served as distractor items in the recognition test. Half of the target items and half of the distractor items were tested in the same context the participant saw at learning; the other half was tested in the other context, that is, a context different from the learning context. The assignment of test items to the two test contexts was randomized by participant.

After completion of the memory test, participants received a computerized version of the Gardner-Monge (Gardner & Monge, 1977) vocabulary test. IBM-compatible personal computers controlled the screen presentation and response collection for the entire experiment as well as the vocabulary test. After completion of the vocabulary test, participants filled out a paper-and-pencil health-and-education questionnaire and were debriefed.

Results and Discussion

Table 1 shows HR, FAR, $d'$s, and context effects for Experiment 1. $d'$ is the discrimination measure based on signal detection theory (SDT; Green & Swets, 1966) and measures an individual’s ability to distinguish old from new items. We chose $d'$ over other available discrimination measures (for a review, see Macmillan & Creelman, 1991) because according to global matching theories, the output from memory in response to recognition test items lies along a continuum of familiarity or match strength. SDT makes this same assumption. However, we also analyzed our data with alternative measures of old–new discrimination, namely, HR – FAR, the discrimination measure of two-high threshold theory (Egan, 1958; Snodgrass & Corwin, 1988) and the
nonparametric discrimination measure $A'$ (Macmillan & Creelman, 1990; Pollack & Norman, 1964). Analyses with these two alternative measures yielded the same patterns of results as did $d'9$ for both our experiments.

For both experiments, we calculated mean $d'9$ scores from individual participant $d'9$s. For the calculation of $d'9$, HR of 1.00 were adjusted to 0.95, and FAR of 0.00 were adjusted to 0.05. This adjustment scheme follows Murnane and colleagues (1999). Another common adjustment scheme is to convert rates of 1.00 and 0.00 to $1 - 1/(2N)$ and $1/(2N)$, respectively (Macmillan & Creelman, 1991). That is, 16 hits (HR = 1.00) are converted to 15.5 hits (HR = 0.97), and 0 false alarms (FAR = 0.00) are converted to 0.5 false alarms (FAR = 0.03). Analysis of our data from both experiments with this latter adjustment scheme yielded the same patterns of results that were found with Murnane and collaborators’ (1999) adjustment scheme. A context effect occurs when HR, FAR, or $d'9$ in same-context testing is larger than HR, FAR, or $d'9$, respectively, in different-context testing. The magnitude of a context effect is calculated by subtracting different-context scores from same-context scores.

Planned comparisons in the form of paired $t$ tests were conducted to test if observed context effects on HR, FAR, and $d'9$ were statistically significant. We chose a conventional alpha level of .05 for all statistical tests reported in this article. According to a priori power analyses, each of the one-tailed $t$ tests we report has a statistical power of .81 to detect a medium effect size ($d' = .5$, see Cohen, 1988, and Appendix, Note 2).

For young adults, both the expected context effect on HR and the expected context effect on FAR were statistically significant, $t(51) = 1.88$, $p = .033$, and $t(51) = 3.31$, $p = .001$, respectively. As predicted, a reliable $d'9$ context effect was not observed. As expected, older adults showed a significant context effect on FAR, $t(51) = 4.17$, $p = .0005$. The context effect on HR was not significant for this age group. According to our predictions, the older group did not show a higher $d'9$ under same-context test conditions compared with different-context test conditions. Related to the (unexpected) absence of a context effect on HR for this age group, $d'9$ was higher in the different-context condition than the same-context condition. Older adults thus showed an unexpected negative context effect.

For young adults, our results replicate those reported by Murnane and Phelps (1993, 1994, 1995) and Murnane and collaborators (1999). Our results, as well as theirs, indicate that young adults use item information ($I$) and associated context ($C$), but not an ensemble of item and context ($E$) in a recognition task with simple visual contexts.

We predicted that with simple contexts, the pattern of context effects would be the same for young and older adults, that is, both groups should show context effects on HR and FAR, and no context effect on $d'9$. This was confirmed by the significant context effect for FAR, and the absence of a positive context effect on $d'9$ in the older age group. The context effect on FAR is predicted by ICE from the use of associated context ($C$). However, if older adults are using associated context, they should also show a context effect on HR, which was not observed. The issue of associated-context use by older adults was further addressed in Experiment 2.

We predicted that in Experiment 1, the second form of context postulated by ICE, namely, the ensemble of item and context, would not be used by either young or older adults. This prediction was confirmed. Use of an ensemble would have resulted in a context effect on $d'9$ (i.e., larger $d'9$ for same than different context), which we did not observe in either age group. These results replicate prior research on young adults regarding the lack of ensemble processing when simple visual context is manipulated (Murnane & Phelps, 1993, 1994, 1995; Murnane et al., 1999) and extend them to older adults.

**Experiment 2**

The purpose of Experiment 2 was two-fold. One purpose was to provide further evidence for the use of associated context by older adults that had been suggested by the context effect on FAR in Experiment 1. The second purpose was to test the hypothesis that older adults have difficulties integrating item and context information into an ensemble. In Experiment 2 we used rich visual contexts (i.e., pictures of scenes) which according to Murnane and collaborators (1999), lead to ensemble formation in young adults. Experiment 2 thus allows a comparison of ensemble formation in young and older adults.

The design and procedure of Experiment 2 were similar to those of Experiment 1. Again, participants performed a relatedness-judgment task on word pairs and, subsequently, received a set of old–new recognition tests. Words were tested either in the same context they had been studied in or in a different context. The major difference between the two experiments is that in Experiment 1 items were presented in simple visual contexts (i.e., combinations of foreground color, background color, and screen location), whereas rich visual contexts (i.e., pictures of scenes) were used in Experiment 2.

We predicted that with rich visual contexts young adults would use both associated context and ensembles and, therefore, show context effects on HR, FAR, and $d'9$, thus replicating the results of Murnane and collaborators (1999). With regard to older adults, we predicted that they would use associated context, but would have difficulties integrating context information with the to-be-remembered item information into an ensemble. Therefore, older adults were predicted to show context effects on HR and FAR, but not on $d'9$. 

### Table 1. Hit Rates (HR), False Alarm Rates (FAR), $d'9$, and Context Effects for Young and Older Adults in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Same context</th>
<th>Different context</th>
<th>Context effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Young adults</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>.86</td>
<td>.11</td>
<td>.83</td>
</tr>
<tr>
<td>FAR</td>
<td>.23</td>
<td>.16</td>
<td>.18</td>
</tr>
<tr>
<td>$d'9$</td>
<td>1.97</td>
<td>.61</td>
<td>1.99</td>
</tr>
<tr>
<td>Older adults</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>.73</td>
<td>.17</td>
<td>.72</td>
</tr>
<tr>
<td>FAR</td>
<td>.29</td>
<td>.17</td>
<td>.22</td>
</tr>
<tr>
<td>$d'9$</td>
<td>1.29</td>
<td>.64</td>
<td>1.52</td>
</tr>
</tbody>
</table>

*Note: n = 52 for both young and older adults.*
Method

Participants.— Fifty-two young adults and 52 older adults participated in Experiment 2. Of the young adults, 40 were undergraduate students at The University of Memphis, and 12 were undergraduate students at The University of North Carolina at Chapel Hill. Twenty of the young participants were men and 32 were women. They were between the ages of 18 and 25 years (mean age 19.8 years), were recruited from introductory psychology courses, and received class credit for their participation. The older adults (17 men and 35 women) were between the ages of 60 and 84 years (mean age 69.4 years), lived in the community, and were recruited through newspaper advertisements and community organizations in the city of Memphis. They received payment for their participation.

All participants were native speakers of English. Mean years of education were 13.8 (SD = 1.1) for young adults and 14.7 (SD = 2.5) for older adults, a significant difference, t(102) = 2.5, p = .014. The mean performance in raw score units on the 30-point Gardner-Monge vocabulary test (Gardner & Monge, 1977) was 12.7 (SD = 3.22) for young adults and 18.8 (SD = 6.17) for older adults, also a significant difference, t(102) = 6.31, p < .0005. In a near-vision test, all participants were able to read text in 7-point size print.

On the health rating scale, 7% of the young adults and 90% of the older adults rated their health as good or excellent. None of the participants rated their health as poor. Health criteria for participation in the experiment were the same as for Experiment 1. The data from 5 older adults were excluded from analyses for the following reasons: (a) One participant did not understand instructions for the relatedness judgment, (b) 1 participant reported that she could not clearly read the words presented in one of the learning contexts, (c) 3 participants did not give a recognition response within the given time limit in four or more of the trials (two in four trials, and one in eight trials), and (d) 1 young participant could not complete the experiment because of technical failure. These participants were replaced by additional participants.

Design.— Like Experiment 1, Experiment 2 also had a 2 × 2 mixed factorial design with age (young vs. older) as a between-subjects variable, and test context (same vs. different) as a within-subjects variable.

Materials.— Stimulus items were the same 96 high-frequency nouns that were also used in Experiment 1 (Francis & Kucera, 1982). They were presented in 22-point-size lower-case black letters. Four colorful pictures of scenes served as rich visual learning contexts: a living room, an airplane, a bus, and a street. These same pictures served as test contexts in same-context conditions. A picture of a classroom served as the context in all different-context test trials. The pictures had been drawn with Microsoft Paintbrush for Windows. An example is given in Figure 1.

Procedure.— Between 1 and 4 participants were tested in each session. Participants were given a consent form and a brief vision test and were directed to individual computer booths. As in Experiment 1, participants were instructed to judge the relatedness of word pairs. A practice list of 8 word pairs preceded a study list of 32 word pairs. The word pairs on the study list were equally divided among four learning contexts such that eight word pairs were presented in each context. Assignment of word pairs to contexts was randomized by participant. The words of each pair appeared on top of each other (with two spaces between the words) within an object in the scene (e.g., on the side of the bus or on the TV in the living room). The word pairs were presented within their picture context for 5 s, after which the sentence “Enter your relatedness judgment” was presented to prompt the relatedness judgment. The next study trial began after participant response or after the prompt had been presented for 5 s.

In the incidental single-item, old–new recognition test, the labeling and assignment of computer keys to response options was the same as in Experiment 1. Again, participants were asked to respond to each test item as fast and as accurately as they could. The recognition test was preceded by eight practice trials. The test word disappeared immediately after the participant responded or after 5 s. The next test trial started after an intertrial interval of 2 s filled with a black, blank screen. No error feedback was provided.

Thirty-two target items and an equal number of distractor items appeared in random order on the recognition test. Target items were randomly drawn from the word pairs that had appeared on the study list with the constraint that half of the target items had appeared as the first member of a studied word pair, whereas the other half of the target items had been the second member of a pair. Distractor items were new items that had not appeared on the study list.

Half of the target items were tested in the same context they had been studied in (same-context test condition); the other half were tested in a new context that the participants had not seen during learning (different-context test condition). Half of the distractor items were assigned randomly and in equal numbers to the four learning contexts (same-context test condition). The remaining distractor items were tested in the new context (different-context test condition). Assignment of items to either same- or different-context test conditions was randomized by participant.

Screen presentation and response collection for the entire experiment was controlled by IBM-compatible personal computers. After completion of the experiment, participants took a computerized version of the Gardner-Monge vocabulary test and a paper-and-pencil health-and-education questionnaire.

Results and Discussion

HR, FAR, d′, and context effects for Experiment 2 are presented in Table 2. Planned comparisons were conducted in the form of paired t tests to test if observed context effects on HR, FAR, and d′ were statistically significant. Young adults showed a significant context effect on HR, t(51) = 4.37, p < .0005. The FAR context effect was in the predicted direction, but did not reach statistical significance, t(51) = 1.13, p = .322. As predicted, a context effect was found in d′, t(51) = 2.26, p = .014, in the younger group. Older adults showed a significant context effect on HR,
Older adults and young adults in our experiment showed context effects on HR and FAR of .03 and .05, respectively. Context effects in this experimental paradigm are usually small, but reliable. The small size of context effects is not surprising given that in both Murnane and collaborators’ (1999) experiments that are comparable to our simple-context experiments in terms of type of context, presentation time, orienting task, and number of item presentations (Murnane & Phelps, 1993, 1994, 1995; Murnane et al., 1999). For these conditions, the weighted mean size of context effects on HR and FAR was .06 and .05, respectively. Context effects in this experimental paradigm are usually small, but reliable. The medium size of context effects are surprising given that in both Murnane and collaborators’ experiments, as well as in our experiment, a small part of the overall context was manipulated. That is, only the appearance of the computer screen was manipulated, whereas virtually all other aspects of the context (room, location in room, presence of experimenter, silence, and so forth) remained constant throughout the experiment.

With regard to \(d'\), Murnane and collaborators (1999) reported context effects from 5 young adult participant groups that saw rich visual contexts at study and at test. These context effects were statistically significant and ranged between .12 and .21 (weighted mean of .16). In our Experiment 2, we found a comparatively large \(d'\) context effect of .24 for young adults. Thus, we created conditions in this experiment that were very conducive to young adults’ use of ensemble information. Yet, under these conditions, our older adults did not show a context effect on \(d'\). The absence of a context effect for older adults with rich contexts corresponds to the absence of context effects that Murnane and collaborators (1999) found with young adults and simple contexts, that is under conditions when young adults do not use an ensemble. The nonsignificant context effect of .03 we found for older adults in Experiment 2 (rich contexts) corresponds very closely to the nonsignificant context effect of .02 found with young adults and simple contexts by Murnane and collaborators (1999) and to the absence of context effects with young adults and simple contexts in six prior experiments by Murnane and Phelps (1993, 1995). Thus, our older adults show a pattern of context effects in Experiment 2 that is predicted when participants do not make use of ensembles. These results provide strong support for the hypothesis that older adults have difficulties integrating item and context information into an ensemble.

It is the integration of item and context that allows the discrimination of old and new items in recognition memory, because an ensemble match increases the global match in response to targets but not to distractors. The lack of integration of item and context by older adults thus offers an explanation for age-related differences in recognition memory performance.

By using a model that distinguishes between different kinds of context processing in recognition tasks, we can

### Table 2. Hit Rates (HR), False Alarm Rates (FAR), \(d'\), and Context Effects for Young and Older Adults in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Same context</th>
<th>Different context</th>
<th>Context effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (M)</td>
<td>SD</td>
<td>Mean (M)</td>
</tr>
<tr>
<td>Young adults</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>.84</td>
<td>.12</td>
<td>.76</td>
</tr>
<tr>
<td>FAR</td>
<td>.19</td>
<td>.17</td>
<td>.16</td>
</tr>
<tr>
<td>(d')</td>
<td>2.06</td>
<td>.69</td>
<td>1.82</td>
</tr>
<tr>
<td>Older adults</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>.72</td>
<td>.20</td>
<td>.68</td>
</tr>
<tr>
<td>FAR</td>
<td>.31</td>
<td>.19</td>
<td>.27</td>
</tr>
<tr>
<td>(d')</td>
<td>1.24</td>
<td>.62</td>
<td>1.21</td>
</tr>
</tbody>
</table>

Note: \(n = 52\) for both young and older adults.

\(t(51) = 2.08, p = .021\), as well as on FAR, \(t(51) = 1.94, p = .029\), confirming our predictions. As also predicted for this age group, there was not a reliable context effect on \(d'\).

The pattern of context effects observed for young adults indicates that they use associated-context information \((C)\) as well as ensemble information \((E)\) when rich visual context is manipulated. This pattern of results replicates that found by Murnane and collaborators. (1999) in two experiments that used rich visual contexts in a total of five experimental conditions. Although the context effect on FAR of .03 that we observed in our experiment did not reach significance, it is similar in size to the context effects on FAR reliably found by Murnane and collaborators (1999). These context effects range from .03 to .07 (weighted mean of .05). Older adults in our experiment showed context effects on HR and FAR, but no context effect on \(d'\). These findings indicate that older adults use item information \((I)\) and associated context \((C)\), but have difficulties in using an ensemble of item and context \((E)\) in the recognition task.

### General Discussion

We derived a complex pattern of predictions from the ICE model and, with very few exceptions, the results of the experiments confirmed those predictions. The results of our research support the hypothesis that older adults encode and use context information in recognition, but do not integrate item and context information as well as young adults do. In two experiments, we investigated age differences in environmental–context effects on recognition and interpreted results based on predictions of the ICE model of recognition memory (Murnane et al., 1999). ICE postulates that three kinds of activation (or match) may play a role in recognition: activation of item information \((I)\), activation of context information that is associated with item information \((C)\), and activation of an ensemble or integration of item and context information \((E)\). Context effects on HR, FAR, and \(d'\) are predicted when \(I, C,\) and \(E\) are activated. Context effects on HR and FAR, but not on \(d'\), are predicted when only \(I\) and \(C\) are activated. In Experiment 1, with simple visual contexts, patterns of context effects showed that young adults used \(I\) and \(C\). Results suggested that older adults may have also used \(I\) and \(C\). This suggestion was confirmed in Experiment 2. In this experiment, which employed rich visual contexts, young adults used \(I, C,\) and \(E\), whereas older adults used \(I\) and \(C\), but not \(E\). In conclusion, older adults encode and use associated context but have difficulties with the integration of item and context information compared with young adults.
show that a general theory that older adults have difficulties in processing contextual information is likely to be incorrect. Our data suggest that older adults use associated context as effectively as young adults do. This finding concurs with conclusions by a number of other researchers (Light et al., 1992; Naveh-Benjamin & Craik, 1995; Vakil et al., 1996). Our data further suggest that the particular problem of older adults is the integration of item and context, a conclusion that was also reached by Park and colleagues (1990) and by A. D. Smith and colleagues (1998) in experiments in which recall paradigms were used.

We have noted that the mechanisms specified in the ICE theory are based on familiarity processes alone, not on recollective processes. We have to consider the possibility that our participants drew on recollective processes in addition to familiarity-based processes when making recognition judgments. In recall tasks, which are based on recollective processes, context effects are found with great consistency regardless of the kinds of contexts used (see S. M. Smith, 1988, for a review). Thus, if recollective processes played a large role in our recognition task, context effects on $d'$ would have been expected in the simple-context experiment (Experiment 1). No such context effects are predicted by the familiarity-based ICE theory. Context effects were not found in either our simple-context experiment or numerous simple-context experiments by Murnane and collaborators (Murnane & Phelps, 1993, 1994, 1995; Murnane et al., 1999). Thus, although we cannot exclude the possibility that participants used recollective in addition to familiarity-based processes in our tasks, the pattern of results with regard to context effects can be explained with familiarity-based processes alone.

Some researchers in the area of memory and aging have suggested that recollective processes are impaired in older adults, whereas familiarity-based processes remain intact (e.g., Jennings & Jacoby, 1993, 1997). The age differences we found in Experiment 2 (rich contexts), however, are explained by a familiarity-based recognition model. We, therefore, conclude that older adults can be disadvantaged in familiarity-based processes if these processes are based on information that requires integration on the part of the participant.

Our results are in line with those of the many studies reporting age differences in source memory. Performance on most source-memory tasks requires the integration of item and context information. So does the formation of an ensemble. Other authors have argued that older adults’ performance on source-memory tasks is impaired, because such tasks require explicit context retrieval in a direct test of context (Light et al., 1992; Naveh-Benjamin & Craik, 1995; Vakil et al., 1996). In our experiments, context was tested in an indirect manner, and age differences did appear that can be explained by differences in ensemble formation. The crucial variable determining if age differences are found is thus not the directness or indirectness of context tests. The crucial variable appears to be the necessity of integration of item and context information.

Although the focus of our study was to reveal age differences in basic memory processes, our findings also have practical implications. If older adults’ problem with episodic-memory tasks lies in a lack of integration of item and context information, the training of such integration might alleviate some of the everyday memory problems encountered by persons of advanced age.

Our study has several limitations that have to be addressed in future research. Although the ICE theory is a general theory of context-dependent recognition and encompasses environmental as well as semantic context, the theory has thus far only been tested with environmental contexts. It remains to be examined if in a paradigm that employs semantic rather than environmental contexts, older adults also show problems of item-context integration. Hess (1984) found context effects for young and older adults when items were presented in semantically related contexts, but context effects for young adults only when items were presented in semantically unrelated contexts. These findings suggest that older adults’ integration of item and context information is facilitated by semantic relatedness, a hypothesis that can be tested with greater precision using the ICE model.

Future research should also address whether conditions can be created under which older adults are more likely than in our present paradigm to integrate item and environmental context information into an ensemble. Last but not least, future research needs to speak to possible explanations for the observed age-related impairment in item-context integration. Our results are in line with a reduced-processing-resource account of cognitive aging (Craik & Byrd, 1982), which postulates that age differences in episodic-memory (and other cognitive tasks) can be attributed to age-related differences in working memory or processing speed (Light, 1991; Park et al., 1996; Salthouse, 1996). The formation of an ensemble at encoding and retrieval may require additional processing resources that may not be available to older adults. Therefore, these adults base their recognition judgments on item (I) and associated context (C) alone. Younger adults might have more processing resources in the form of higher working-memory capacity or higher processing speed, allowing them to form and use an ensemble in addition to item and associated-context information. It remains to be determined if individual differences in older adults’ ensemble formation can be explained by differences in measures of processing resources.

Our research offers an explanation for age-related differences in recognition memory derived from ICE, a model of recognition memory that emphasizes integrative and nonintegrative processing of contextual information. As indexed by their use of associated context, older adults engaged in nonintegrative processing of contextual information as effectively as young adults did. However, as indicated by their deficit in ensemble formation compared with that of young adults, older adults showed difficulties in self-initiated integration of item and context information. Older adults appear to have a specific, rather than a general, deficit in processing contextual information. Their difficulty with ensemble formation may explain why older adults tend to show poorer performance on recognition-memory tasks than young adults.

Acknowledgments

The research reported in this article was supported by a Junior Faculty Development Award, a University Research Council Grant, and a Cogni-
tive Science Research Seed Grant from The University of North Carolina at Chapel Hill, as well as a Project Grant from the Institute on Aging of The University of North Carolina System. Preparation of this article was supported in part by Research Grant AG17456 from the National Institute on Aging. Parts of this research were presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, November 1997.

We thank Andreas Enneking, Destiny Shellhammer, Alicia Hill, Jill Zukerman, Erica Snider, Jamie Newman, Victoria Pomeroy, and Richard Carroll for assistance with participant recruitment and data collection, and appreciate the help of the Chatham County and Chapel Hill Senior Centers. We thank Kevin Murnane for many discussions of the ICE model.

Address correspondence to Ute J. Bayen, Department of Psychology, The University of North Carolina, Davie Hall, CB#3270, Chapel Hill, NC 27599, E-mail: ubayen@unc.edu

References


Appendix

Note 1. ICE can predict present or absent context effects on $d'$ when only item and associated-context information contribute to the global match. Which prediction is made depends on properties of the activation function used to combine item and associated-context match (i.e., whether or not the effect of associated-context match on the global match depends on the magnitude of the item match with which it is combined) and on the magnitude of match-mismatch differences achieved in a particular experimental paradigm. See Murnane and collaborators (1999) for a more thorough discussion of this issue and evidence that associated context is not likely to produce $d'$ context effects in the experimental paradigms used in this research.

Note 2. The power analyses reported in this article were computed by means of the GPOWER program (Faul & Erdfelder, 1992; see Erdfelder, Faul, & Buchner, 1996).