Universität Potsdam

Laura A. Thompson, Reinhold Kliegl

Adult age effects of plausibility on memory : the role of time constraints during encoding

first published in:
Journal of Experimental Psychology 17 (1991) 3, S. 542-555,
ISSN 1939-1285, DOI 10.1037/0278-7393.17.3.542

Adult Age Effects of Plausibility on Memory: The Role of Time Constraints During Encoding

Laura A. Thompson and Reinhold Kliegl
Max Planck Institute for Human Development and Education, Berlin, Federal Republic of Germany

We investigated the role of training-induced knowledge schemas and encoding time on adult age differences in recall. High-plausible (schema coherent) words were recalled better than low-plausible (schema discrepant) words in both age groups. This difference was larger for old adults than for young adults for presentation times ranging from 3 s to 11 s per word. After equating participants in overall recall (i.e., at 50% correct) by dynamic adjustment of presentation time, old adults again showed a stronger plausibility effect than young adults when recall was above criterion. In a second experiment with self-paced encoding, old adults used more time than young adults only for low-plausible pairs, yet they still remembered fewer of them. In a third experiment, both age groups preferred to imagine high- rather than low-plausible words, but this effect was more pronounced in old adults. The results indicate that, compared with young adults, old adults find it particularly difficult to form elaborative mental images of schema-discrepant information under a wide variety of time constraints during encoding. Results are discussed in relation to explanations based on age-related mental slowing.

In previous adult memory-training studies, young adults initially recalled more words than old adults and also benefited more from instruction and practice with elaborative encoding (e.g., Kliegl, Smith, & Baltes, 1989; Rose & Yesavage, 1983). For example, Kliegl et al. (1989) reported considerably more improvement by young adults compared with old adults in the level of serial-word recall attained after training in the method of loci mnemonic technique. The method of loci involves forming mental images of the to-be-remembered words in connection with schemas of an overlearned sequence of landmarks. One factor that could affect the ease or difficulty of connecting thoughts at any age, but especially in old age, concerns the relationship between the items to be remembered and their respective mnemonic pegs. Relevant evidence suggests that old adults are not as proficient as the young in generating word associations between pieces of unrelated information (e.g., Rabinowitz, Craik, & Ackerman, 1982b). If this assumption holds true for image-like processing, then old adults should experience comparably more difficulty than young adults when performing encoding and recoding operations connecting items to unrelated landmarks, as in the combination Botanical Garden–computer. Alternatively, the words that may be easiest to recode into interactive images are those that have a plausible relation to preexisting knowledge structures of familiar landmarks, such as Botanical Garden–rose. In our terminology then, a plausible or schema-coherent relation exists if the to-be-remembered item is often experienced by an individual at a particular landmark or if world knowledge would lead one to expect a likely association.

There are two main goals of this article. First, we want to establish that adult age differences in generating schema-coherent versus schema-discrepant relations can be obtained in a context where task-relevant schemas are induced by mnemonic training. Second, we want to examine the effect of encoding time for generating schema-coherent and schema-discrepant relations. Possibly the age-differential effects reported in previous research had less to do with the aspect of information integration per se but more with the general difficulty in generating such relations given the constraints implied by mental slowing in advanced age (Salthouse, 1980; Waugh & Barr, 1980).

Age Effects of Relatedness on Recall

Three lines of research support the proposition that, compared with young adults, old adults would be expected to encode schema-coherent information far more easily than

This study was supported in part by a postdoctoral fellowship from the Max Planck Institute to Laura A. Thompson. Portions of this article were written while Laura A. Thompson was at New Mexico State University. This research is part of the project "Expertise and Cognitive Aging" co-directed by Reinhold Kliegl and Paul B. Baltes. We are grateful to Annette Rentz and Werner Scholtysik for coordinating the research and for technical assistance. We also thank the following individuals for research assistance: Gregor Caregnato, Jackie Durfee, and the late Sabine Edler. We are saddened by the loss of our friend and colleague. Several others helped us improve our study through collegial discussion or written commentary, including Paul B. Baltes, Steve Cornelius, Anders Ericsson, Ulman Lindenberger, Timothy A. Salthouse, and Jacqui Smith. We also thank Janice Keenan and two anonymous reviewers for valuable comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to Laura A. Thompson, who is now at the Psychology Department, Box 3452, New Mexico State University, Las Cruces, New Mexico 88003; or to Reinhold Kliegl, Max Planck Institute for Human Development and Education, Lentzeallee 94, D-1000 Berlin 33, Federal Republic of Germany.

schema-discrepant information. In pictorial memory tasks, older adults have been found to show an advantage for remembering related information relative to young adults (Azmitia, Merriman, & Perlmutter, 1987; Waddell & Rogoff, 1981). For example, Azmitia et al. (1987) presented young adults, old adults, and children hand-drawn scenes containing high- and low-expectancy items, which they were to remember for recall and recognition tests. Interestingly, old adults and children, but not young adults, showed better recall for high- compared with low-expectancy items.

A second basis of support comes from a comparison of two well-known studies of adult age differences in text comprehension. Belmore (1981) and Cohen (1979) arrived at very different conclusions concerning adult age differences in drawing inferences during text processing. Belmore (1981) found no age-differential decline in inferential processing, whereas Cohen (1979) found that young adults were superior to the old in forming accurate inferences. As Zacks and Hasher (1988) noted, the reason for the conflicting results could be that Belmore's (1981) materials incorporated more well-learned preexisting relations than those used by Cohen (1979). (For a similar argument, see Reder, Wible, & Martin, 1986.)

Finally, a third line of research showed that age differences in paired-associate learning vary as a function of preexperimental associative strength between stimulus and response elements. In addition to a general age difference in learning and easier learning of related items, age differences in the number of trials to criterion are larger if pairs are not related or are only weakly related (Kausler & Lair, 1966; Ross, 1968; Zaretsky & Halberstam, 1968). Similar results were obtained in cued-recall tasks varying the relatedness of cue and target (Rabinowitz, 1986; Rabinowitz et al., 1982b; Shaps & Nilsson, 1980). However, Rabinowitz, Ackerman, Craik, and Hinchley (1982a) did not obtain any significant interactions with age in a cued-recall task varying relatedness of cue–target pairs and imagery instruction. There were strong effects of imagery instruction and relatedness, and imagery instruction reduced the effect of relatedness, but the significant age difference in cued recall in favor of young adults was constant across the experimental conditions.

The issue of a general age difference in memory as a function of schema relatedness was the focus of the first and second experiments of the present study. Old and young individuals were trained in the method of loci technique. They learned to recall a set of West Berlin landmarks in a specified order, and learned to generate interactive visual images of landmark–noun combinations, although only words were presented during encoding. That is, they were taught to rely on a mental set of mnemonic pegs for encoding and retrieval of a list of words. In the critical experimental sessions, one third of the words constituted highly plausible relationships with the mnemonic pegs. For different participants, the same words were also presented at serial positions, which formed less plausible relations with the location schemas. Thus, we examined the effect of plausibility by pairing the to-be-encoded words with both related and unrelated mental-landmark schemas.

Mental Slowing as an Explanation of the Effect of Relatedness

Why might older adults be expected to experience difficulty performing elaborative encoding processes on schema-discrepant information? Whether or not one obtains an effect of relatedness might depend on the time available to encode and elaborate information. Specifically, if older adults are generally slower in performing mental operations (e.g., Birren, 1974; Salthouse, 1985), this should exacerbate their difficulty integrating schema-discrepant compared with schema-coherent information when the encoding conditions do not allow enough time to form a high-quality memory trace (Salthouse, 1980; Waugh & Barr, 1980). Some evidence from paired-associate learning paradigms is consistent with this position, reporting smaller age differences as the length of the anticipation interval is increased (Arenberg, 1965; Canestrari, 1963; Monge & Hultsch, 1971; Treat & Reese, 1976; for a review see Witte, 1975). However, in studies using cued or free recall, or those increasing the study time, no age-differential effects of presentation times were found; in fact, some studies report even larger benefits for young adults (Craik & Rabinowitz, 1985; Kliegl et al., 1989; Monge & Hultsch, 1971; Rabinowitz, 1989; Rankin & Hinrichs, 1983; Smith, 1976; Treat & Reese, 1976). These results are not necessarily a problem for the mental-slowing hypothesis because recall could depend on the number of elaborations or rehearsals, and young adults may be able to complete a greater number of such elaborations in a given time than old adults.

In any case, if the age difference in memory for schema-coherent compared with schema-discrepant items is maintained over both short and long encoding times, this would be consistent with a memory-specific age difference in encoding schema-discrepant information. In contrast, a change in the age difference of the plausibility effect as a function of available encoding time suggests that age differences in the ease of encoding schema-discrepant material might be contingent on available processing time, which in turn might be a consequence of general mental slowing.

In the present study, we investigated possible trade-offs in the type of material best remembered with variations in task difficulty by having participants perform the task under different presentation-time conditions. In Experiment 1, in addition to presenting word lists at various fixed presentation times, we also determined functionally equivalent presentation times for each individual. Specifically, we determined for each individual the presentation time at which performance was maintained at a 50% correct level of serial recall across lists. If such criterion-referenced testing eliminates the expected interaction between age and degree of relatedness using fixed presentation-time conditions, this would strengthen the interpretation that age-differential effects of relatedness are an epiphenomenon of older adults' slower processing speeds.

Equating individual levels of recall by criterion-referenced adjustment of presentation times serves two additional purposes. First, if participants recall 15 of 30 words after a single presentation of the list, the probability that they did not use the mnemonic strategy is very small. This procedure thus ensures that, at least at a global level, young and old adults engage in comparable cognitive processes. Second, the criterion-referenced testing procedure represents a direct manipulation of encoding with strong implications for standard scores of recall accuracy. Specifically, Kliegl and Lindenberger (1989) obtained (negative) correlations above .70 between serial-recall accuracy with fixed presentation times and criterion-referenced encoding times within groups of young and old adults. The high correlations were consistent with the interpretation that the fastest encoders in the criterion-referenced testing condition had better opportunities for performing a greater amount of elaborative processing steps in the fixed presentation-time conditions, resulting in higher quality and hence more memorable images. Moreover, age differences in recall accuracy were completely accounted for by individual differences in criterion-referenced encoding times. Therefore, there is some face validity to the claim that the speed of generating images is a critical factor for within-groups and between-groups memory performance of adult age groups.

In Experiment 2, the interaction between processing speed and the age-differential effect of relatedness was further scrutinized. Participants in this experiment were given an unlimited amount of time to forge an association between a given landmark and the item to be remembered. Under self-paced encoding conditions, both young and old adults were predicted to use relatively more time encoding schema-discrepant compared with schema-coherent items; however, this difference should be greater for old adults if forming elaborative images of schema-discrepant information is in fact more difficult for old adults than it is for young adults.

Experiment 1

Method

Participants

Seventeen old adults (M = 73 years; range = 65–85 years; 13 women and 4 men) and 19 young adults (M = 24.6 years; range = 22–29 years; 14 women and 5 men) were recruited by newspaper ads and by advertisement at the Free University of Berlin to participate in the study. In addition to the final group of 36 participants, 2 other old adults began the study but could not complete it because of illness, whereas 1 old and 2 young adults were not asked to return after intelligence quotient (IQ) testing revealed low IQ scores. Participants were paid DM 20 (i.e., about $12) for each of 14 experimental sessions.

Young adults had completed 13 years of formal schooling and had attended an average of 5 years at the university. The average number of years the adult group attended school was 11.3. Five of the old adults had attended college. Thus, the old adult group had less formal education. On a 5-point subjective health scale (1 = *very poor*, 5 = *very good*), the old and young did not differ significantly. Both groups rated their health as "good," an average of 4.1 on the rating scale.

Pretraining Intelligence and Memory Functioning

Eight subscales of the German version of the Wechsler Adult Intelligence Scale (HAWIE; Wechsler, 1956) were administered to all participants before the experimental phase. Post hoc t tests revealed that the young adults scored significantly higher than the old on three performance IQ subscales ($p < .05/8 = .007$). However, the two age groups did not score significantly different from each other on the verbal subscales. This pattern of intelligence scores is consistent with the pattern of normal aging. Seven memory subscales from the Nuremburg Aging Inventory (NAI; Oswald & Fleischmann, 1986) were also administered. These data are reported in a later article focusing on age-differential relations between pre-experimental memory ability and mnemonic training gains.

Overview of Experimental Sessions

Individuals participated in 14 laboratory sessions. In Session 1, participants were given the HAWIE and NAI tests. Session 2 involved a pretraining assessment of serial-word recall with different presentation times. In Sessions 3 and 4, participants received instruction in the method of loci mnemonic technique. Baseline and final assessments of serial-word recall using the method of loci (Sessions 5 and 12) preceded and followed training. Here presentation times were the same for all members of an age group, but also included a different set of times across age groups. In six criterion-referenced testing sessions (Sessions 6 to 11), presentation times were dynamically adjusted to each individual's 50% level of correct cued recall. Session 13 was used to collect plausibility ratings of the memory stimuli. Session 14 was Experiment 3.

Apparatus and Materials

Apple IIe computers displayed the words in standard Apple 40-column font on an Apple IIe monitor. Thirty well-known West Berlin landmarks served as the mnemonic pegs. Three types of words served as items to be remembered. Items that were likely to be encountered at, or highly associated with, a given landmark will be called high-plausible items. For example, a sphinx is highly associated with the Egyptian museum. A set of 18 high-plausible associates was generated for each landmark, and 9 of these words appeared at their respective location in one set of materials (Form A). The other 9 words in this set, referred to as low-plausible words, appeared at a single different landmark. This partner landmark was chosen at random respecting the following three constraints: The partner could not be a neighbor in the mental map or a landmark of the same type (e.g., another castle), and two landmarks could not exchange both high-plausible and low-plausible items. In Form B, this assignment was reversed, so that the items that appeared at their low-plausible locations in Form A appeared at high-plausible locations in Form B, and vice versa. By placing words at both high-plausible and low-plausible locations, we controlled for the potential confound of item and plausibility effects. The third type of words, referred to as random items, were filler items and bore no special relationship to their mnemonic pegs. Random words were the same across Form A and Form B lists. One half of the participants within the two age groups received Form A, and the other half received Form B.

The dimension of plausibility was assessed by questionnaire after the conclusion of the experiment. All of the participants were asked for plausibility ratings of high-plausible and low-plausible words from the study. To keep the number of ratings within a 1-hr completion period, random words were not rated. They rated the question "How likely would a person experience or think of (the word) at this particular (West Berlin) location?" on a 6-point Likert scale. Both groups gave significantly higher ratings to the words when rated at their high-plausible locations (old-5.35; young-5.26) compared with their low-plausible locations (old-2.27; young-2.40). The age differences in ratings for high-plausible or low-plausible words were not significant.

Design and Procedure

Pretraining serial-recall assessment (Session 2). Participants took part in a 1-hr serial-recall test before mnemonic instruction. The computer presented each word for 11 s in the first list, then for 8 s, 5 s, 3 s, 2 s, and 1 s in the following lists. No cues appeared at either the encoding or the retrieval phase. Participants wrote their responses for each list on paper after being instructed to attempt to place them in their correct position in the list. They were not under pressure to respond quickly. Responses were scored with a strict criterion: Both word and absolute position in the list had to be correct.

Instruction in the method of loci (Sessions 3 and 4). Instruction in the method of loci occurred in two sessions. In the first session, the tutor described the method of loci according to the recommendations of Bower (1970). Participants were shown color photographs of the 30 landmarks and were asked to build their own visual images for each location. They then practiced using the technique to recall three 10-word lists. Successive sets of 10 landmarks were used for this practice. Landmark cues were visible during encoding and recall. Words were read by the tutor. The tutor evaluated the memory images and emphasized the need to create interactive, dynamic images or thoughts. Participants were told that they had to overlearn the list of 30 landmarks for the next session. In the following session, participants were asked to recite the map of 30 landmarks. Training proceeded only if participants could recite the map in the correct order within 90 s; time was also devoted to rehearsing the list in the laboratory. Practice was identical to the previous session, except that no location/landmark cues were used during encoding; participants had to call on their mental map to perform the task.

Baseline and final assessments of mnemonic skill (Sessions 5 and 12). In the baseline and final-assessment sessions, word lists were presented by computer. One third of the words in each 30-word list were high plausible, one third were low plausible, and one third were randomly selected from a 1,560-word pool of concrete German nouns. Words never appeared twice within the entire set of word lists in the experiment for a given person. In the young-adult group, presentation times were 11 s, 8 s, 5 s, 3 s, 2 s, and 1 s per word from the first to the sixth list, respectively. There were six lists presented in each session. For old adults, the presentation times used were 20 s, 15 s, 11 s, 8 s, 5 s, and 3 s per word. Thus, across age groups, common presentation times were 11 s, 8 s, 5 s, and 3 s per word. These times were chosen to minimize floor effects in old adults and ceiling effects in young adults. Moreover, in anticipation of large age differences, we wanted to assess older adults under a comparatively easy condition and young adults under a comparatively difficult condition to compare the groups at similar levels of performance. Participants wrote their responses on a sheet of paper containing landmark cues after presentation of each list. A maximum of 10 min was allotted for the recall of a list; this time was never fully exhausted. These sessions were conducted in age-homogeneous groups of 3 to 5 persons. Each individual sat at a computer monitor out of view of other participants.

Criterion-referenced testing (Sessions 6–11). There were five 30-item lists to be memorized during each of six criterion-referenced testing sessions. In Sessions 7, 9, and 11, word lists were constructed of 10 high-plausible, 10 low-plausible, and 10 random words. Participants in the old-adult group began with a 13-s-per-word presentation time, whereas the young-adult group began with a 7-s presentation time. If the individual achieved a 50% level of correct serial recall or better, presentation time was shorter on the following list. Likewise, a performance worse than 50% correct resulted in a longer presentation time on the next list. At rates higher than 7 s, the adjustment occurred in 2-s increments. Between 7 s and 3 s, the incremental steps moved up or down 1 s; and below 3 s, the adjustments occurred in 0.5-s incremental steps. At retrieval, the computer displayed the landmark cues in a random order, and allowed a maximum of 20 s for each response to be told to the tutor, who then typed the response into the computer. After recall of a list, the computer displayed the landmarks, the correct words, and the person's responses in three columns. Criterion-referenced testing sessions were individualized.

Results and Discussion

Results are presented in four sections: (a) serial-recall performance immediately after instruction in the method of loci (baseline) and after criterion-referenced testing (final); (b) correct recall and presentation times during criterion-referenced testing; (c) above- and below-criterion performance; and (d) error analysis. In all analyses, the criterion level for statistical significance was $p < .05$.

Baseline and Final-Assessment Sessions

Immediately after instruction in the method of loci and after the six criterion-referenced testing sessions, participants' mnemonic skill was assessed at six different presentation times. Corresponding means and standard deviations, also broken down by age group and type of relation, are shown in Table 1. These data were analyzed in two steps. In the first step, we report the analysis on overall level of cued recall and age-differential training gains. In the second step, we focus on the effect of plausibility on cued recall.

Overall level of recall. A repeated-measures analysis of variance (ANOVA) was performed using the variables age group (2), time of assessment (2), and presentation time (4 [the times common to the two age groups]). Young adults recalled more words than old adults, $F(1, 34) = 84.2$, $MS_e = 153.7$. The amount of improvement across time was significant, $F(1, 34) = 55.2$, $MS_e = 26.9$, but young adults improved more than old adults, $F(1, 34) = 7.9$, $MS_e = 26.9$. This interaction is graphically portrayed in Figure 1, which plots the mean correct recall for the two age groups as a function of time of assessment, collapsing across the 11-s through 3-s presentation times and word types. Also included in Figure 1 are the averages for old and young during the pretest session before instruction in the mnemonic technique.

Recall was higher with long presentation times, $F(1, 34) = 77.7$, $MS_e = 18.1$. Presentation time interacted with age group, $F(1, 34) = 10.6$, $MS_e = 18.1$. This interaction was further qualified by time of assessment, $F(1, 34) = 3.9$, $MS_e = 12.4$, for a three-way interaction. At final assessment, age differ-

Table 1
Serial-Word Recall as a Function of Time of Assessment, Type of Relation, Presentation Time, and Age Group

Type of relation	Young: High plausible	Young: Low plausible	Young: Random	Young: Total	Old: High plausible	Old: Low plausible	Old: Random	Old: Total
				Baseline assessment				
Presentation time (s)								
20					4.1 ± 3.2	3.2 ± 3.2	3.1 ± 3.4	10.5 ± 9.5
15					4.9 ± 2.7	3.0 ± 2.3	3.0 ± 3.6	10.9 ± 8.3
11	8.6 ± 1.4	7.5 ± 2.2	7.6 ± 2.2	23.7 ± 5.2	3.8 ± 2.5	1.9 ± 2.7	2.7 ± 2.6	8.3 ± 6.7
8	8.2 ± 2.8	7.4 ± 3.1	7.2 ± 2.6	22.7 ± 7.9	3.8 ± 3.2	1.9 ± 1.8	1.5 ± 2.4	7.1 ± 6.6
5	5.7 ± 2.6	5.1 ± 2.3	4.1 ± 2.6	14.8 ± 7.0	2.0 ± 1.8	1.2 ± 1.4	2.0 ± 1.5	5.2 ± 4.0
3	4.4 ± 2.6	3.2 ± 2.2	1.9 ± 2.4	9.5 ± 6.4	1.7 ± 1.3	0.8 ± 0.6	0.8 ± 1.0	3.3 ± 2.4
2	2.2 ± 2.3	1.2 ± 1.3	2.3 ± 1.9	5.7 ± 4.6				
1	1.3 ± 1.1	1.2 ± 0.9	0.8 ± 0.8	3.4 ± 2.1				
				Final assessment				
Presentation time (s)								
20					7.1 ± 2.7	6.1 ± 3.3	6.1 ± 2.8	19.3 ± 8.0
15					6.0 ± 2.8	5.0 ± 3.2	4.1 ± 3.2	15.1 ± 8.6
11	9.6 ± 0.8	9.4 ± 1.2	9.3 ± 1.0	28.4 ± 2.5	5.6 ± 2.5	3.7 ± 2.8	4.1 ± 2.7	13.4 ± 7.6
8	9.5 ± 0.7	9.0 ± 0.8	9.3 ± 1.0	27.7 ± 1.7	4.2 ± 2.6	3.6 ± 2.9	2.4 ± 2.7	10.2 ± 7.7
5	8.4 ± 1.7	7.5 ± 1.7	6.8 ± 2.0	22.7 ± 4.6	3.0 ± 2.5	1.8 ± 2.4	1.8 ± 1.9	6.6 ± 6.2
3	7.0 ± 2.3	5.6 ± 2.5	4.4 ± 2.3	17.1 ± 6.5	2.7 ± 2.5	1.4 ± 1.5	1.1 ± 1.5	5.1 ± 5.2
2	6.3 ± 2.2	4.3 ± 2.2	3.6 ± 1.8	14.2 ± 5.6				
1	2.6 ± 2.0	1.7 ± 1.7	2.0 ± 1.7	6.3 ± 5.0				

Note. Values are means ± standard deviations. Maximum total score was 30. Only words that were recalled in the original serial position were scored as correct answers.

ences were larger for the long (i.e., 11 s and 8 s) than for the short presentation times (i.e., 5 s and 3 s). At baseline, group differences did not vary as a function of presentation time. These results are in partial agreement with those of Kliegl, Smith, and Baltes (1990), who reported larger practice gains for young adults than for old adults after instruction in the method of loci for the 5-s condition but not for shorter (3 s, 1 s) and longer (10 s, 15 s, 20 s) presentation times. In Figure 1, serial-word recall before mnemonic instruction is displayed as well. Instruction in the method of loci increased serial-word recall in young and old adults, but differentially more so for young adults, a result also reported by Kliegl et al. (1989).

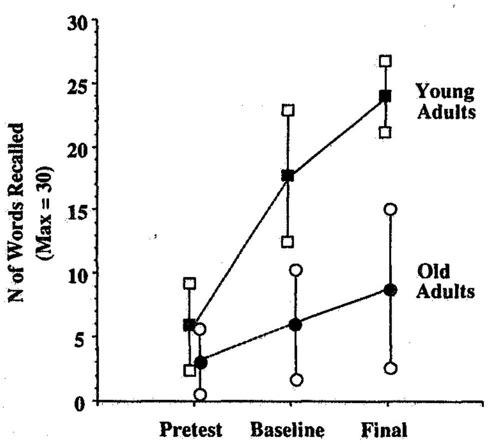

Figure 1. Average serial recall for old and young adults in Experiment 1. (All scores averaged across three word types. Scores are averaged over the 11-s, 8-s, 5-s, and 3-s, presentation times for baseline and final assessments. Bars indicate 1 *SD*.)

Plausibility effect: raw scores. Across participants, the same set of words was presented at different list positions, forming both high-plausible and low-plausible relations with their respective mental landmarks. Hence, the following analysis addressed the main question of this article: Was the plausibility effect greater for older adults than for younger adults? A 2 × 2 × 4 × 2 (Age Group × Time of Assessment × Presentation Time × Type of Relation) repeated measures ANOVA was performed on the average number of words recalled for high-plausible and low-plausible landmark–noun combinations.[1] High-plausible relations were recalled more often than relations of low plausibility, $F(1, 34) = 81.2$, $MS_e = 2.0$. Most important, the interaction between type of rela-

[1] The analyses focus on the contrast between high-plausible and low-plausible landmark–noun relations because they involved the same set of nouns. Descriptive statistics associated with random landmark–noun relations are included in the tables.

tion and age group was marginally significant, $F(1, 34) = 3.8$, $MS_e = 2.0$, $p < .06$, and in the predicted direction: The difference between recall of high-plausible relations was larger for old adults than for young adults. Type of relation did not interact with time of assessment or presentation time. (Other effects were as described previously for the complete set of 30 words.)

Plausibility effect: ratio scores. Because overall level of recall was far lower for old than for young adults, raw scores for the two age groups do not contain the same level of informativeness. Ratio scores may be more informative because they take into account the number of high-plausible and low-plausible words recalled given the total number of words recalled. Summing across presentation times and across time of assessment, old adults recalled words 2.23 times more often in relation to high-plausible mental landmarks than they recalled these same words in relation to low-plausible landmarks. The corresponding ratio for young adults was 1.67. In a 2 × 2 (Age Group × Time of Assessment) repeated measures ANOVA of ratios, this difference between age groups was highly significant, $F(1, 34) = 32.3$, $MS_e = 0.63$. Moreover, the advantage of plausible relations decreased from baseline to final assessment for both groups, $F(1, 34) = 6.2$, $MS_e = 0.62$.

Preliminary summary. With fixed presentation times, old adults tended to recall more high-plausible than low-plausible words than did young adults. This plausibility effect was marginally significant for raw scores and highly significant for ratio scores. These results are in agreement with the research reviewed earlier here suggesting that, during encoding, old adults rely more on preexisting knowledge structures than do young adults. Note that the plausibility effect was obtained at rather low levels of recall, indicating that participants were using their mental landmarks even with very fast presentation times. We did not obtain a three-way interaction among type of relation, presentation time, and age group. It is possible, however, that this null result was due to ceiling and floor constraints. For slow rates (11 s, 8 s), young adults were close to ceiling at least in the final assessment. For fast rates (5 s, 3 s), old adults' very low level of recall may have masked a stronger plausibility effect. Criterion-referenced adjustment of presentation times avoids both floor and ceiling effects and might provide a clearer picture in this respect.

Criterion-Referenced Testing

Correct recall and presentation times. During criterion-referenced testing sessions, presentation times were dynamically adjusted to each individual's 50% level of correct recall. The observed pattern of means averaged across the five lists in a session are shown separately for the two age groups in the left panel of Figure 2. The figure shows that the young-adult group reached a 50% level of performance in the third session. Old adults' 50% performance occurred in the second session. There were no significant differences involving age group and session after Session 2. Thus, recall performance was functionally equivalent beginning in Session 3 of criterion-referenced testing.

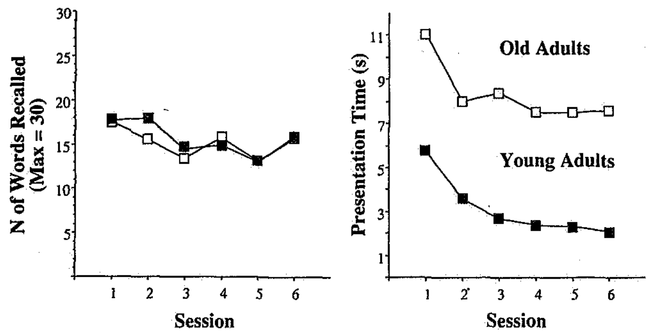

Figure 2. Left panel: Average correct serial recall as a function of age group and criterion-referenced testing sessions; right panel: average presentation time as a function of age group and criterion-referenced testing sessions. (Presentation times were individually determined and depended on the recall level in preceding lists. Each data point represents the average of five lists.)

The right panel of Figure 2 displays the associated pattern of presentation times, averaged across individuals within age group and across the five lists within each of the six criterion-referenced testing sessions. Note that encoding times for old adults are more than three times longer than for young adults. In both age groups, the rates show a marked decline between the first and second training sessions and a tapering off by the third session. The data were analyzed in a 2 × 4 (Age Group × Session) ANOVA with repeated measures on the session factor. The analysis revealed significantly longer presentation times for old adults than for young adults, $F(1, 34) = 44.44$, $MS_e = 22.90$. Participants also became significantly faster across these four training sessions, $F(3, 102) = 3.00$, $MS_e = 1.24$; however, this effect did not interact with age group.

Plausibility effect: raw scores. In Sessions 4 and 6, at equivalent levels of overall recall, two thirds of the items in each list were high-plausible and low-plausible landmark-noun relations. These data were analyzed with a 2 × 2 × 2 (Age Group × Time of Assessment × Type of Relation) repeated measures ANOVA on raw scores for high- and low-plausible words. High-plausible relations were recalled significantly better than low-plausible relations, $F(1, 34) = 312.8$, $MS_e = 0.37$. No other main effects or interactions were significant (all Fs < 1.2).

Plausibility effect: ratio scores. Plausibility ratios for Sessions 4 and 6 were also computed. Averaging across sessions, the ratios were 1.42 for old adults and 1.40 for young adults, which was not a significant difference. In contrast to the results from the fixed presentation-time analyses, no differences in the strength of the plausibility effect were obtained when comparing old and young adults at equivalent recall levels.

Above- and Below-Criterion Recall in Criterion-Referenced Testing

The lack of an age-differential plausibility effect with criterion-referenced presentation times implied, at first glance, that age differences in plausibility could be considered a mere epiphenomenon of criterion-referenced encoding time. It is important to realize, however, that equal recall across lists arose as a consequence of averaging over conditions in which the task can be performed rather well and conditions that may not allow one to carry out the elaborative processing steps connected with the method of loci. The overall absence of an age-differential plausibility effect could be the consequence of age-differential responsiveness to such challenges of the skill.

To put this result to another test, we evaluated the plausibility effect as a function of whether the 50% recall criterion had been met or not. Across Sessions 4 and 6, on average, more than 15 of 30 words had been recalled in 5.4 lists for old adults and 5.3 lists for young adults. Thus, as expected by the criterion-referenced test procedure, in each age group about half of the 10 lists resulted in above- and below-criterion performance, respectively. In a 2 × 2 × 2 (Age Group × Type of Relation × Level of Recall) repeated measures ANOVA, there were significant effects of level of recall, $F(1, 34) = 300.6$, $MS_e = 1.1$, a significant interaction between age group and level of recall, $F(1, 34) = 5.3$, $MS_e = 1.1$, a significant effect of type of relation, $F(1, 34) = 309.0$, $MS_e = 0.4$, and a significant interaction involving the three factors, $F(1, 34) = 6.5$, $MS_e = 0.3$. Means and standard deviations are displayed in Table 2.

To specify the source of the three-way interaction more precisely, two post hoc ANOVAs were performed, one for above-criterion lists and one for below-criterion lists. For lists with a high level of recall (see Table 2), a significant interaction was obtained between age group and type of relation, $F(1, 34) = 4.7$, $MS_e = 0.4$: Old adults were more affected by the plausibility manipulation than were young adults; they recalled significantly fewer low-plausible relations of above-criterion lists than young adults, $t(34) = -2.4$. The corresponding interaction in an ANOVA for lists with a low level of recall was not significant, $F(1, 34) = 1.4$, $MS_e = 0.2$; there was, however, a tendency for old adults to perform better than young adults irrespective of plausibility, $F(1, 34) = 3.7$, $MS_e = 1.6$, $p < .07$. In summary, in lists with above-criterion recall, we still observed a significant age-differential plausibility effect. This result is critical because, for these lists, we can be confident that the mnemonic device was used by old and young adults.

Table 2
Recall for Above- and Below-Criterion Lists During Criterion-Referenced Testing Sessions as a Function of Type of Relation and Age

Type of plausible relation	Above criterion[a]		Below criterion[b]	
	High	Low	High	Low
Old	7.9 ± 0.5	5.9 ± 0.8	5.1 ± 0.9	3.4 ± 0.9
Young	7.8 ± 0.8	6.5 ± 0.7	4.7 ± 1.2	2.7 ± 0.8

Note. Values represent means ± standard deviations. Ten is the maximum number of words for each condition.
[a] Above criterion is greater than 15 items correctly recalled in a list.
[b] Below criterion is 15 or fewer correctly recalled items.

Error Analysis

Age-differential effects of plausibility were obtained under fixed and criterion-referenced presentation times. An alternative explanation for age-differential tendencies in using preexisting knowledge schemas would be age-differential guessing strategies. For example, old adults may be more willing to guess a plausible word under fixed presentation-time conditions.

To address this question, all false responses were rated according to whether they were plausible at the landmark or not. Across all lists administered after instruction (12 baseline and final lists and 30 criterion-referenced testing lists) 1,916 errors were counted. The agreement between two raters was 95% across errors. Disagreements were mostly due to one of the raters applying a stricter criterion of plausibility (21 vs. 91 errors in the disagreement cells). The raters agreed on 245 errors as being plausible for the landmark.

Plausible and nonplausible errors were categorized according to whether they occurred (a) in baseline and final assessment session lists, (b) in the first two criterion-referenced testing sessions for which age differences in recall were obtained, or (c) in the final four criterion-referenced testing sessions with equivalent recall in both age groups. Means and standard deviations for individual sums of errors across lists within these categories are displayed in Table 3 for both age groups along with t statistics for the difference between them. The results indicate that old adults made more plausible false responses when recall was not equivalent, but that there were no age differences when recall had been equated by criterion-referenced testing; there were also no age differences for above- and below-criterion lists. Moreover, old and young adults did not differ significantly in other false responses.[2]

The error analyses suggested that the age-differential plausibility effect observed under fixed presentation-time conditions could have been due to older adults' tendency to produce more plausible false responses than young adults. We computed a plausible error ratio (i.e., number of plausible errors divided by number of nonplausible errors) analogous to the plausibility ratio for recall. Using this error ratio as a covariate, we still obtained a significant age difference ($p < .01$) for the

[2] As an alternative method of determining plausible and nonplausible errors, we had the computer check all false responses against the set of 18 words selected to serve as plausible items for the landmark at which the error occurred. This analysis resulted in the same pattern of significance as the data based on the ratings; as expected, means for plausible errors were only slightly lower, indicating that most of the plausible false responses originated from previously presented plausible items.

Table 3
Plausible and Nonplausible False Responses as a Function of Type of Test and Age

Variable	Baseline/ final tests	Criterion-referenced testing (Sessions 1, 2)	Criterion-referenced testing (Sessions 3–6)
Plausible errors			
Old	5.7 ± 4.0	1.4 ± 1.1	2.2 ± 2.1
Young	2.8 ± 2.8	0.3 ± 0.8	1.5 ± 1.8
$t(34)$	2.54*	3.31**	1.16
Other errors			
Old	27.7 ± 12.1	6.0 ± 4.4	16.4 ± 12.9
Young	23.9 ± 12.9	8.7 ± 10.8	10.5 ± 12.3
$t(34)$	0.90	−0.98	1.40

Note. Values are means ± standard deviations.
* $p < .05$. ** $p < .01$; other errors do not include omissions.

plausibility recall ratio. In summary, although there was evidence for age-differential tendencies to give plausible false responses under conditions with unequal levels of recall, this response bias was not strong and general enough to account for the age-differential plausibility effect.

Experiment 2

As expected, with fixed presentation times, old adults in Experiment 1 recalled proportionately more high-plausible than low-plausible words than did young adults. Alternatively, with presentation times individually adjusted to yield 50% recall, the interaction was obtained for lists with 50% recall or better. An alternative approach to validating the age-differential plausibility effect is to examine self-paced encoding times. In Experiment 2, participants were allowed to study at their own pace. If the integration of low-plausible landmark–noun relations is more difficult for old adults than for young adults, old adults should spend relatively more time encoding low-plausible than high-plausible pairs compared with young adults.

Method

Participants

Fifteen old adults (M = 71.9 years; range = 65–87 years; 11 women and 4 men) and 15 young adults (M = 25.3 years; range = 20–30 years; 8 women and 7 men) participated in the study. As in Experiment 1, participants were paid in DM about $12 for each of four experimental sessions. None of the individuals had previously participated in a memory-training study.

Young adults had currently completed an average of 4.2 years of college in addition to 13 years of previous formal education. For the old-adult group, the average number of years of schooling was 11.9; 6 participants had taken university courses. Thus, the old-adult group had less formal education. On the subjective health scale, old adults rated themselves at 3.7 and young adults rated themselves at 4.3. Unlike Experiment 1, the age difference in subjective health ratings was significant, $t(28) = 2.2$, $p < .05$. Digit–symbol substitution was administered as a marker of general intellectual ability; raw scores were 4.13 for old adults and 58.2 for young adults, $t(28) = 5.2$, $p < .01$. Old adults' digit–symbol score was somewhat lower than typically obtained for samples of healthy, old adults. However, across experiments, t tests comparing young and old groups revealed no significant differences for age, digit–symbol substitution scores, health status, and education. Thus, although the present sample of old adults appeared to be somewhat lower in level of cognitive functioning, it still was comparable to the sample of old adults participating in the first experiment.

Apparatus and Materials

As in Experiment 1, Apple IIe computers presented the stimuli and collected response times during encoding. The experiment comprised four sessions. The three lists for Session 1 were composed of random concrete nouns. Word lists used in Sessions 2, 3, and 4 were the same as those used in the baseline and final sessions of Experiment 1. That is, there was a total of 12 lists, each composed of 10 items from the following three categories: high-plausible, low-plausible, and random words. The low-plausible and high-plausible words were identical, but they were presented at landmarks forming high- and low-plausible landmark–word relationships. Random words were the same words presented at the same landmark positions in Forms A and B. As before, Forms A and B were present to half of the participants in each age group.

Design and Procedure

In Session 1, participants completed a demographic questionnaire and the digit–symbol substitution test. Then they were asked to encode and recall a list of 30 words with cues present during encoding and recall as described later. After this pretest, they learned to visualize the West Berlin landmarks and to form interactive images. They did not, however, learn to recall the landmarks in a specific order. Two practice lists were administered after instruction.

Sessions 2, 3, and 4 were experimental. During these sessions, participants encoded the words in each of four lists at their own pace. The words and their associated landmark cues were presented in two successive lines in the middle of the computer screen. Participants could take as long as 60 s to encode each list item. When they were finished with an item, they pressed the space bar on the computer keyboard, which signaled the computer to display the next landmark and word in the list. The landmarks were presented in the same order for all individuals and on every list.

At retrieval, the landmark cues were presented in a randomized order. Randomization was different for each individual and each list. This procedure eliminated the opportunity for using placement in the list as a cue for recall. Participants would read the landmark, and then they attempted to think of the word that was associated with the landmark in that list. The next step was to write the word down on a sheet of paper. If the word could not be recalled, subjects indicated this by drawing a dash in the appropriate position in the list. In either case, the computer displayed the next cue on the keyboard after participants signaled their readiness for the next item by depressing the space bar or after 30 s had elapsed. Four 30-item lists were given during each experimental session. Sessions were conduced in small, age-homogeneous groups.

Results and Discussion

Recall Performance

There were large differences between age groups in recall performance. Averaging across Sessions 2 to 4 and lists, young adults recalled 8.9 (SD = 0.5) high-plausible and 8.6 (0.9) low-plausible pairs. The corresponding values for old adults were 6.1 (2.2) and 4.5 (2.6), respectively. Young adults' performance was close to ceiling, as indicated in the standard deviations (maximum score was 10 for each word type).

Number of recalled high- and low-plausible words were analyzed with a 2 × 3 × 2 (Age Group × Session × Type of Relation) ANOVA specifying session and type of relation as within-subject factors (see Footnote 1). The analysis yielded significant effects of age group, $F(1, 28) = 29.8$, $MS_e = 18.0$, and type of relation $F(1, 28) = 61.3$, $MS_e = 0.8$, and a significant interaction between these two factors, $F(1, 28) = 24.3$, $MS_e = 0.8$. The plausibility effect was significant in both age groups. The smaller difference between recall of low-plausible and high-plausible pairs in young adults cannot be interpreted unambiguously because of ceiling problems. For example, 6 young participants recalled all 10 high-plausible items in at least 7 of the 12 lists administered. Finally, recall increased across sessions, $F(2, 56) = 4.9$, $MS_e = 0.8$. Adding together only the high-plausible and low-plausible items recalled, the following means were obtained: 13.7, 13.7, and 14.6 for Sessions 2, 3, and 4, respectively.

For the old adults, interindividual differences in recall were stable across the three experimental sessions. As shown above the diagonal in Table 4, correlations ranged from .84 to .96. Young adults' values were substantially lower, again an indication of the ceiling effect.

Self-Paced Encoding

Self-paced encoding times give a direct measure of age differences in the difficulty of forming images of plausible and implausible memory material. Table 5 displays means and standard deviations for encoding times broken down by age group, type of landmark–noun relation, and recall status (recalled vs. not recalled). The table shows that the difference in encoding times for high- and low-plausible words was larger for old adults regardless of whether or not the encoded items were correctly recalled.

Encoding times for high- and low-plausible pairs were analyzed with a 2 × 2 × 2 (Age Group × Type of Relation × Recall Status) ANOVA, with type of relation and recall status as within-subjects factors. Because of the age differences in recall, the encoding times for recalled items were more reliably measured in young adults compared with old adults, whereas the reverse was true for items that were not recalled.[3] High-plausible pairs were encoded faster than low-plausible pairs (12.2 s vs. 15.2 s), $F(1, 28) = 42.6$, $MS_e = 7.0$. There was also a significant interaction between age group and type of relation, $F(1, 28) = 4.2$, $MS_e = 7.0$: The age difference in encoding times was smaller for high-plausible (old-12.5 s; young-11.8 s) than for low-plausible pairs (old-16.7 s; young-14.0 s). Finally, there was also a significant interaction between type of relation and recall status, $F(1, 28) = 14.7$, $MS_e = 1.8$: There was a more pronounced difference in encoding times for low-plausible items that were recalled and those that were not (15.6 s vs. 14.8 s) compared with the corresponding times for high-plausible pairs (12.0 s vs. 12.4 s).

Table 4
Stability of Encoding Times (Below Diagonal) and Recall Performance (Above Diagonal) Across Three Sessions as a Function of Type of Relation and Age

	High plausible			Low plausible			Random		
Group	1	2	3	1	2	3	1	2	3
Old									
1	1.00	0.88	0.85	1.00	0.85	0.84	1.00	0.90	0.93
2	0.84	1.00	0.84	0.80	1.00	0.95	0.90	1.00	0.96
3	0.79	0.82	1.00	0.82	0.78	1.00	0.83	0.85	1.00
Young									
1	1.00	0.51	0.32	1.00	0.85	0.43	1.00	0.53	0.47
2	0.89	1.00	0.32	0.88	1.00	0.37	0.88	1.00	0.84
3	0.88	0.94	1.00	0.84	0.95	1.00	0.89	0.94	1.00

Note. Numerals 1, 2, and 3 indicate experimental session.

Effects of Plausibility and Item Encoding Time on Recall

Plausibility and encoding time were hypothesized to be important determinants of recall. The effects of these variables on recall can be tested simultaneously at an individual level. For each participant, a logistic regression analysis was performed on the 360 items presented with recall status as criterion and plausibility and encoding time as predictors.[4] We used the constrained nonlinear regression (CNLR) module of SPSS-X, which also provides standard errors for coefficients by means of a bootstrapping procedure (SPSS Inc., 1988). The coefficients for the 15 young and 15 old adults were analyzed with an ANOVA containing age group (2) as the between-subjects factor and coefficients (4) as the within-subjects factors. Simple effects were tested for each coefficient (i.e., whether a coefficient was significant overall and whether there was a significant difference between groups).

Recall level. Group differences in recall after controlling for encoding time and plausibility are reflected in the regression intercept. This coefficient was significant overall, $F(1, 28) = 11.2$, and it was significantly higher for young adults than for old adults, $F(1, 28) = 33.6$, $MS_e = 1.4$. For old adults alone, the coefficient was not significant. At an individual level, this coefficient was significantly positive for 4 old and 12 young adults.

[3] Five young participants had sessions in which their recall was perfect. For this reason, we did not include session as a factor in the ANOVA. When only correct encoding times were analyzed, there was a significant decrease across sessions; the speedup was larger for low-plausible than for high-plausible pairs. There were no age-differential effects associated with session.

[4] Initially a multiplicative interaction term between plausibility and encoding time was also used as a predictor. Because of recall ceilings for plausible items, this analysis could not be performed for some participants. When the interaction coefficient could be estimated, it was never significant.

Table 5
Self-Paced Encoding Times as a Function of Type of Relation, Recall Status, and Age

	High plausible		Low plausible		Random	
Group	Recalled	Not recalled	Recalled	Not recalled	Recalled	Not recalled
Old	12.5 ± 3.5	12.6 ± 3.6	17.6 ± 5.1	15.7 ± 4.7	16.7 ± 4.9	17.2 ± 5.7
Young	11.5 ± 7.3	12.2 ± 7.4	14.1 ± 8.1	13.9 ± 7.7	14.5 ± 8.0	15.8 ± 10.0

Note. Values are means ± standard deviations.

Plausibility. The effect of plausibility was significant overall, $F(1, 28) = 75.7$, and there was a significant difference between age groups, $F(1, 28) = 20.5$, $MS_e = 0.2$, indicating that plausibility was more predictive of recall in old than in young adults. At an individual level, this coefficient was significantly positive for 12 old and 3 young adults.

Encoding time. Intraindividually determined effects of encoding time were not significant for young or old adults, $F(1, 28) = 1.2$ for overall effect; $F < 1$ for the age difference. At an individual level, the effect was significantly positive for 3 old adults. Thus, spending more time on an item increased the chance of recalling it after statistically controlling for the plausibility of the item for only 3 participants.

To summarize, the results of Experiment 2 showed that when adults used as much encoding time as they believed they needed to successfully retrieve information from an elaborated image, both age groups took more time to encode schema-discrepant information, yet they were still less likely to correctly recall it compared with schema-coherent information. More important, old adults spent more time encoding schema-discrepant (low plausible) items compared with young adults. Age comparisons based on logistic regression analyses for individual subjects are complicated by ceiling effects for young adults. For old adults, the effect of plausibility could be clearly established at this level even if encoding time was statistically controlled. In the same context, it became clear that intraindividual variability in encoding time did not predict recall after plausibility was partialed out. Thus, it appears that schema coherence is more important for later recall than time spent elaborating the relation.

Experiment 3

In Experiment 3, we examined whether people believe that extra time or effort is required to form elaborative images of schema-discrepant information. An awareness of the relative ease of encoding high-plausible relations should have impact on the amount of time or effort people use to form elaborative images if there are strong time constraints. In the present experiment, we gave the participants of Experiment 1 the choice-of-image task. They were asked which of three items, presented on a computer screen below one of the 30 landmarks, "would be easiest to imagine for correct later recall" at the specified landmark. We expected that participants from both age groups would indicate a strong preference for words that were plausibly related to their mnemonic pegs compared with less plausible relations. Moreover, old adults should show an even stronger preference to imagine schema-coherent information than young adults, given that old adults required comparatively more time to encode schema-discrepant information.

Method

Participants

All of the 34 participants in the choice-of-image task (16 old; 18 young) had completed the sessions of Experiment 1. One of the old adults from Experiment 1 could not be in Experiment 3 because of illness, and technical malfunctioning necessitated the exclusion of 1 young adult's data.

Apparatus, Design, and Materials

The same Apple IIe computers used in the previous two experiments were used in presenting the stimuli and collecting responses, including reaction times to the nearest millisecond (Poltrock & Foltz, 1982). In the choice-of-image task, each trial contained one of the 30 Berlin landmarks and three concrete German nouns (choices). For 270 of the 360 total trials, one of the choices formed a high-plausible, one a low-plausible, and one a random landmark–noun pair. In the remaining 90 trials, all three choices were random associates, selected from the pool of words used in the previous experiments. These random-item sets were included to steer participants away from a strategy of searching for the item that was associated with the landmark without spending time attempting to visualize the items. The choices making up the high-plausible and low-plausible pairs came from the same set of materials as did the memory materials from Experiment 1: Form A and Form B. However, if a participant received Form A materials in Experiment 1, that same person received Form B materials in Experiment 3. So, for example, if in Experiment 1 an individual saw rose paired with Botanical Garden on one trial (high plausible), radar paired with Botanical Garden on another trial (low plausible), and kite at Teufelsberg Park (high plausible) on yet another trial, the same individual, in Experiment 3, would see rose at the International Congress Center (low plausible), radar at Teufelsberg Park (high plausible) and kite at the Botanical Garden (low plausible). Hence, the high-plausible and low-plausible words had been seen by everyone in Experiment 1, but they appeared in connection with different landmarks in Experiment 3. To control for familiarity of word types, the random items also came from lists in the previous experiment. The order of the landmarks was consistent with the sequence of the mental map. Whether or not a particular trial consisted of random or mixed landmark–noun pairings was determined randomly.

Procedure

Participants were instructed to carefully read the information on the screen and then to choose which item (Numeral 1, 2, or 3 on the keypad) that they thought they would have the least difficulty remembering after making their elaborative image for it. They were told to perform the task at a comfortable pace, but to avoid undue delay because reaction times were being collected. Stimuli were presented in three 90-trial blocks, with a 5-min break between blocks. Participants were tested in age-homogeneous groups of 3 to 5 persons in a 1-hr session.

Results and Discussion

The responses were computed for those trials where all three types of landmark–noun relationships were presented as choices, and the results are displayed in Table 6 for the two age groups. The top part of the table shows that both groups' stated preference was to imagine the items forming high-plausible landmark–noun pairs over the low-plausible words. In addition, it can be seen that old adults selected the high-plausible items more often than did the young adults. These observations were confirmed by statistical analyses of high-plausible and low-plausible choices. In a 2 × 2 repeated measures ANOVA, significant main effects were found for age, $F(1, 32) = 9.5$, $MS_e = 340.8$, and choice (ignoring random choices), $F(1, 32) = 133.2$, $MS_e = 2{,}378.6$. The old adults' choices from these two categories were higher overall (because they gave fewer random choices), and individuals made more high-plausible than low-plausible choices. The Age × Choice interaction was significant, $F(1, 32) = 10.3$, $MS_e = 2{,}378.6$, because of the larger difference in older adults' selection of high-plausible than low-plausible items compared with young adults.

Response times are shown in Table 6. A 2 × 2 repeated measures ANOVA revealed a significant effect for type of relation, $F(2, 64) = 8.26$, $MS_e = 1{,}116{,}795.2$. There was no significant effect for age group, nor did the interaction reach statistical significance. Post-hoc t tests revealed that choices for high-plausible relations were made more quickly than for both low-plausible, $t(33) = 3.21$, and random $t(33) = 4.30$, relations; however, responses to random and low-plausible relations were not significantly different. The rather long response times indicate that both groups made their selections carefully. The lack of an interaction with age shows that the groups were comparable in terms of the amount of thought they gave to this task.

The results from the choice-of-image task showed that both old and young adults preferred to imagine high-plausible items. However, old people indicated they would choose the high-plausible items for imaging at a particular location more frequently than the young. Thus, the age difference in selection preferences corroborates the finding from Experiments 1 and 2 that, relative to young adults, old adults find it much easier to form elaborative images of items that are associated with location schemas.

Table 6
Number of Items Chosen and Corresponding Response Times in the Choice-of-Image Task as a Function of Type of Relation and Age

Variable	High plausible	Low plausible	Random
No. of items			
Old	210 ± 51	35 ± 26	26 ± 26
Young	158 ± 42	59 ± 21	53 ± 26
Response time (s)			
Old	4.6 ± 1.4	5.3 ± 1.8	5.8 ± 2.5
Young	4.2 ± 0.9	5.1 ± 2.4	4.8 ± 1.4

Note. Values are mean ± standard deviation.

General Discussion

A consistent theme has emerged from our study: During the construction and recall of elaborative images connecting familiar mental landmarks and nouns, old adults experience comparatively more difficulty than young adults recalling schema-discrepant relations. This pattern of results was obtained under a wide variety of manipulations of encoding times. In Experiment 1, at fixed presentation times ranging between 3 s and 11 s per word, old adults recalled proportionately more high-plausible words relative to low-plausible nouns compared with young adults. During criterion-referenced testing sessions, encoding times were adjusted as a function of each individual's recall on the last list. On lists where individuals exhibited greater than 50% recall, old adults again showed a stronger effect of plausibility than young adults. Under the self-paced encoding conditions of Experiment 2, old adults used proportionately more time forming low-plausible landmark–noun images than did young adults. Moreover, individually based logistic regression analyses indicated that plausibility was a strong determinant of recall for most old adults but only for a few young adults; in contrast, intraindividual variability in encoding time was not critical after statistically controlling for plausibility. Finally, Experiment 3 revealed that old adults believed that encoding words that are related to mental landmarks is more likely to result in higher recall than encoding words that are not related. Thus, old adults demonstrated their difficulty forming and recalling novel relations: (a) by their greater tendency to recall high-plausible nouns, (b) by the greater amount of time they needed to encode them irrespective of later recall, and (c) in their stronger preference, relative to young adults, to steer away from these types of images.

How do our results compare with those of other studies on adult age differences in semantic processing? At a task-specific level, the results indicate an age difference in the ability to integrate novel information into available knowledge schemas. Old adults find it more difficult to integrate information that is inconsistent with their landmark schemas. The role of knowledge schemas in adult age has received attention recently in work on text comprehension. Across studies, the results are not clear-cut. Age differences in memory were reported to be invariant across manipulations of schema availability in three experiments by Arbuckle, Vanderleck, Harsany, and Lapidus (1990). Alternatively, Hess (1985), who manipulated the amount of context provided by scripts, found smaller age differences in memory for typical compared with atypical actions. Furthermore, studies mentioned initially here, such as paired-associated learning studies (Kausler & Lair, 1966; Ross, 1968; Zaretsky & Halberstam, 1968) and cued-recall studies varying the relatedness of cue and target (Rabinowitz, 1986; Rabinowitz et al., 1982b; Shaps & Nilsson, 1980), showed age-differential effects of relatedness on recall, and are thus consistent with our findings with image-like processing; an exception in this respect is the study by Rabinowitz et al. (1982a), who found no interaction between age and relatedness on recall. In sum, the results from our study are consistent with the bulk of the evidence indicating a selective deficit for older adults in remembering implausible relations.

In contrast to the task-specific level of interpretation in terms of knowledge integration, our study was designed to test a possible alternative, specifically that age-differential plausibility effects are a consequence of general age-related resource limitations (Craik, 1983) or general cognitive slowing (Salthouse, 1985). From this perspective, the more difficult or complex-task conditions would be expected to lead to larger age differences. In the present study, the lower recall of low-plausible compared with high-plausible pairs reflects the fact that low-plausible pairs are more difficult in general; consequently, age differences associated with them are larger. We manipulated general task difficulty by treating available encoding time (i.e., self-paced, fixed, and criterion-referenced presentation times) as the critical resource. Thus, we tested the hypothesis that general resource deficits were responsible for the age-differential plausibility effect by determining whether the amount of available encoding time would interact with age and plausibility. The dominant result was that the plausibility effect was maintained across all of these manipulations with one exception: Criterion-referenced adjustment of presentation times across age groups resulted in the absence of an Age × Plausibility interaction on lists when recall fell below 50%, that is, under conditions where it was uncertain whether participants could still use the mnemonic device. Moreover, logistic regression analyses of data from Experiment 2 revealed that old adults' intraindividual variability in encoding time did not predict recall after statistically controlling for plausibility. Thus, the age-differential plausibility effect appears to reflect a phenomenon that is not easily accounted for by general accounts of resource limitations, at least if these are operationalized in terms of available encoding time. Moreover, the error analyses indicated that most likely it is not just a reflection of age-differential response bias.

Although criterion-referenced testing did not eliminate the age-differential plausibility effect, the interindividual differences in memory accuracy assessed in baseline and final test sessions were captured quite well by this procedure. Specifically, the correlation between accuracy in recall with fixed presentation times and the amount of time needed during encoding to maintain 50% correct recall were −.75 for young adults and −.79 for old adults; individuals with the shortest encoding times exhibited the highest recall under standard assessment. Thus, at the level of interindividual differences in memory ability, irrespective of the plausibility manipulation, the data were consistent with the expectation that slow, elaborative processing may be a determinant of poor recall. At this point, however, we cannot rule out the alternative explanation that people with good retrieval strategies do not need to generate high-quality memory traces and, therefore, need less time for elaboration than people with inefficient or deficient retrieval strategies.

In conducting experiments of adult age differences in cognitive processing, although rarely done, it is also important to design experiments that extend across several sessions. Age-comparative research is susceptible to the troublesome influence of age-differential performance factors. Previous studies have documented a reluctance on the part of old adults to even attempt to generate elaborations that stray too far from past experience. For instance, old adults resisted using bizarre imagery in a paired-associates task (Poon & Walsh-Sweeney, 1981) and rejected a high proportion of paired-associate items as too odd to form relations (Hulicka & Grossman, 1967). In the context of the method of loci task, young adults would be expected to be less reluctant to generate an image involving words such as computer for the mnemonic peg Botanical Garden than old adults. One may reasonably expect that the influence of nonexperimental variables is particularly pronounced in initial sessions. As old adults gain more experience with the task, the reluctance to generate "odd" relations may decline because attempting to encode all words increases the opportunities for recalling more words; in this respect, old adults should become more and more like young adults, which should also limit the influence of possible cohort effects. Consequently, if the effects of relatedness decrease across sessions, this constitutes strong evidence for the role of factors unrelated to memory processes per se.

There was no evidence in our study that practice had an age-differential impact on the plausibility effect. Rather there was a remarkable stability of effects across a very lengthy schedule of experimental sessions. Both age groups showed a more marked improvement for low-plausible compared with high-plausible information. For example, in both age groups, the advantage of plausible items in recall decreased from baseline to final assessments with fixed presentation times. Moreover, under self-paced encoding conditions, the speedup in encoding times was larger for low-plausible compared with high-plausible information; yet again the amount of improvement was the same for young-adult and old-adult groups. All persons who completed the 14 sessions constituting Experiments 1 and 3 were very motivated and eager to participate as indicated, for example, by their adherence to session schedules. In sum, there was no indication that the old adults differed from the young adults in their willingness to put effort into encoding schema-discrepant information, and there was also no indication that encoding schema-discrepant information became differentially more automatic for older adults with greater practice.

Training gains were quite compatible with theoretical expectations about age-differential limits in developmental reserve capacity or learning potential (Baltes, 1987; Kliegl & Baltes, 1987). Replicating the results of a previous study (Kliegl et al., 1989, 1990), mnemonic training was found to accentuate the differences in recall between young and old adults. This pattern of results is counter to the intuition that cognitive training should reduce the amount of internal resources required for cognitive processing and, consequently, lead to a reduction in age differences (e.g., Rabinowitz et al., 1982a). The results are, however, consistent with resource-deficit accounts of cognitive aging (Craik, 1983; Salthouse, 1985), if one allows that mnemonic training increases the degree of self-initiated processing both in young adults and old adults, and that older adults are limited in the amount of resources they can allocate.

A final comment pertains to the criterion-referenced assessment of cognitive processes. Our expectation had been that criterion-referenced testing would equate young adults and old adults not only in overall level of recall but also in susceptibility to plausibility. The fact that young adults and old adults continued to differ in this respect for lists with above-criterion recall argues for the specificity of age deficits in knowledge integration. Irrespective of the specific outcome of the present study, the strategy of equating overall quality of performance across age or other nonexperimental groups by means of manipulating some critical resource such as encoding time could be used in a wide variety of task domains. The method may be particularly useful to delineate task-specific explanations from explanations in terms of a general task difficulty and to control for speed–accuracy trade-offs. Furthermore, an important agenda for future research is to specify the conditions under which general age-differential performance limitations are translated into task-specific performance deficits. In this respect, criterion-referenced testing in combination with cognitive training could be a useful tool as an experimental supplement to statistical control techniques.

References

Arbuckle, T. Y., Vanderleck, V. F., Harsany, M., & Lapidus, S. (1990). Adult age differences in memory in relation to availability and accessability of knowledge-based schemas. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 303–315.

Arenberg, D. (1965). Anticipation interval and age differences in verbal learning. *Journal of Abnormal Psychology, 70,* 419–425.

Azmitia, M., Merriman, W. E., & Perlmutter, M. (1987). A life-span study of the interaction of selectivity and knowledge in memory. *Child Development, 58,* 276–281.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics of growth and decline. *Developmental Psychology, 23,* 611–623.

Belmore, S. H. (1981). Age-related changes in processing explicit and implicit language. *Journal of Gerontology, 36,* 316–322.

Birren, J. E. (1974). Translation in gerontology: From lab to life: Psychophysiology and speed of response. *American Psychologist, 29,* 808–815.

Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist, 58,* 496–510.

Canestrari, R. E. (1963). Paced and self-paced learning. *Journal of Gerontology, 18,* 165–168.

Cohen, G. (1979). Language comprehension in old age. *Cognitive Psychology, 11,* 412–429.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London, B302,* 341–359.

Craik, F. I. M., & Rabinowitz, J. (1985). The effects of presentation rate and encoding task on age-related memory deficits. *Journal of Gerontology, 40,* 309–315.

Hess, T. M. (1985). Aging and context influences on recognition memory for typical and atypical script actions. *Developmental Psychology, 21,* 1139–1151.

Hulicka, I. M., & Grossman, J. L. (1967). Age-group comparisons for the use of mediators in paired-associate learning. *Journal of Gerontology, 22,* 46–51.

Kausler, D. H., & Lair, C. V. (1966). Associative strength and paired associate learning in elderly subjects. *Journal of Gerontology, 21,* 278–280.

Kliegl, R., & Baltes, P. B. (1987). Theory-guided analysis of development and aging mechanisms through testing-the-limits and research on expertise. In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 95–119), Norwood, NJ: Ablex.

Kliegl, R., & Lindenberger, U. (1989). *A mathematical model of proactive interference in cued recall: Localizing adult age differences in memory functions.* Manuscript submitted for publication.

Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology, 25,* 247–256.

Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Developmental Psychology, 26,* 894–904.

Monge, R. H., & Hultsch, D. F. (1971). Paired-associate learning as a function of adult age and the length of the anticipation and inspection intervals. *Journal of Gerontology, 26,* 157–162.

Oswald, W. D., & Fleischmann, U. M. (1986). *Nürnberger-Alters-Inventar (Nuremberg Aging Inventory).* Erlangen: Universitat Erlangen-Nurnberg.

Poltrock, S. E., & Foltz, G. S. (1982). An experimental psychology laboratory system for the Apple II microcomputer. *Behavior Research Methods and Instrumentation, 14,* 103–108.

Poon, L. W., & Walsh-Sweeney, L. (1981). Effects of bizarre or interacting imagery on learning and retrieval in the aged. *Experimental Aging Research, 7,* 65–70.

Rabinowitz, J. C. (1986). Priming in episodic memory. *Journal of Gerontology, 41,* 204–213.

Rabinowitz, J. C. (1989). Age deficits in recall under optimal study conditions. *Psychology and Aging, 4,* 378–380.

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I. M., & Hinchley, J. L. (1982a). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology, 37,* 688–695.

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

Rankin, J. L., & Hinrichs, J. V. (1983). Age, presentation rate, and the effectiveness of structural and semantic recall cues. *Journal of Gerontology, 38,* 593–596.

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

Rose, T. L., & Yesavage, J. A. (1983). Differential effects of a list-learning mnemonic in three age groups. *Gerontology, 29,* 293–298.

Ross, E. (1968). Effects of challenging and supportive instructions on verbal learning in older persons. *Journal of Educational Psychology, 59,* 261–266.

Salthouse, T. A. (1980). Age and memory: Strategies for localizing the loss. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging* (pp. 47–65). Hillsdale, NJ: Erlbaum.

Salthouse, T. A. (1985). *A theory of cognitive aging.* Amsterdam: North Holland.

Shaps, L. P., & Nilsson, L. G. (1980). Encoding and retrieval operations in relation to age. *Developmental Psychology, 16,* 636–643.

Smith, A. D. (1976). Aging and the total presentation time hypothesis. *Developmental Psychology, 12,* 87–88.

SPSS Inc. (1988). *SPSS-X user's guide.* Chicago, Illinois: SPSS Inc.

Treat, N. J., & Reese, H. W. (1976). Age, pacing, and imagery in paired-associate learning. *Developmental Psychology, 12,* 119–124.

Waddell, K. J., & Rogoff, G. (1981). Effect of contextual organization of spatial memory of middle-aged and older women. *Developmental Psychology, 17,* 878–885.

Waugh, N. C., & Barr, R. (1980). Memory and mental tempo. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.). *New directions in memory and aging* (pp. 251–260). Hillsdale, NJ: Erlbaum.

Wechsler, D. (1956). *Die Messung der Intelligenz Erwachsener. Textband zum Hamburg-Wechsler-Intelligenztest fur Erwachsene (HAWIE).* Huber: Vern und Stuttgart.

Witte, K. L. (1975). Paired-associate learning in young and elderly adults as related to presentation rate. *Psychological Bulletin, 82,* 975–985.

Zacks, R. T., & Hasher, L. (1988). Capacity theory and the processing of inferences. In L. L. Light & D. M. Burke (Eds.), *Language, memory, and aging* (pp. 154–170). New York: Cambridge University Press.

Zaretsky, H., & Halberstam, J. L. (1968). Age differences in paired-associate learning. *Journal of Gerontology, 23,* 165–168.