Journal of Experimental Psychology:
Learning, Memory, and Cognition
1986, Vol. 12, No. 1, 72–81

0278-7393/86/$00.75

Differential Memory Changes With Age: Exact Retrieval Versus Plausible Inference

Lynne M. Reder
Carnegie-Mellon University

Cynthia Wible
Johns Hopkins University

John Martin
ITT Advanced Technology Center, Shelton, Connecticut

Elderly and college-age subjects were compared in two experiments, one involving episodic memory and one involving semantic memory. Responses were generally slower for older subjects; however, in some conditions, older subjects were as good as or better than young subjects, even in terms of response time. The results suggest that older subjects have no difficulty with memory tasks that do not require exact memory-matches or careful inspection of retrieved propositions. It is argued that careful inspection is a much more costly process for older adults than it is for young adults, but that plausibility judgments and feature overlap processes are equally easy for both age groups. The suggestion is made that older subjects also tend to modify their performance in ways that would minimize the detrimental effects of forgetting specific facts.

A salient attribute associated with increasing age, from young adulthood to retirement and beyond, is a decrement in memory performance (see Birren & Schaie, 1977; Hunt & Hertzog, 1981; Poon, Fozard, Cermak, Arenberg, & Thompson, 1980, for extensive reviews). Although there is almost universal agreement that memory performance declines with age in certain situations, there is less consensus as to the nature of the decline or the cause. Researchers have argued as to whether the deficits are primarily encoding-based, storage-based or retrieval-based, or some combination of these, and whether deficits should be thought of as a reduction in capacity, loss of processing power, loss of processing speed, or loss of mental flexibility (see Craik, 1977, Kausler, 1982, and Salthouse, 1982, for extensive reviews).

Arguments have been made that the more difficult the memory task, the greater the disparity in performance between older and younger adults (e.g., Craik, 1968; Laurence, 1967). On this, there is relative consensus. It is not surprising, for example, that the disparity in performance between younger and older adults would be greater for recall than for recognition (e.g., Botwinick & Storandt, 1974; Erber, 1974; Schonfield & Robertson, 1966). Recall is less likely to suffer from ceiling effects and requires more processes than does recognition, so there are more processes or stages for an age advantage to emerge. More surprising are claims that the disparity in performance increases with more "meaningful" tasks. Recall of word strings is better for all subjects as the strings more closely approximate English; however, the age difference in performance grows with the closer approximations to English (Craik & Masani, 1967). Similarly, Botwinick and Storandt (1974) found that age differences were less pronounced for recall of nonsense-sentences than sensible sentences.

Other research (e.g., Walsh & Baldwin, 1977; Walsh, Baldwin, & Finkle, 1980) seems to suggest that when subjects are asked to recall the gist of a sentence rather than provide verbatim recall, the age differences in memory are much smaller. Mandel and Johnson (1984) found that older adults' recall of stories was both quantitatively and qualitatively similar to that of young adults. They argued that discourse processes do not undergo substantial change during adulthood "because of the greater automaticity of many of the processes involved in comprehension and recall of stories." Indeed, they cite 10 other studies that also show similar qualitative patterns in story recall across various age groups.

Automaticity has been advanced as an explanation, or at least as a construct, in understanding which cognitive skills do not decline with age. The distinction between automatic and controlled processes was first introduced by Posner and Snyder (1975). When applied to understanding the problems of aging, it has been called the attention deficit hypothesis (Hunt & Hertzog, 1981). There is evidence that those processes that require little conscious attention (and are, therefore, called *automatic*) are less likely to degrade with age (e.g., Attig & Hasher, 1980; Cohen & Faulkner, 1983; Craik & Byrd, 1982; Hasher & Zacks, 1979; Kausler & Hakami, 1982; McCormack, 1981; Rabinowitz, Craik, & Ackerman, 1982; Zacks, 19_2). An example of an automatic task investigated by Hasher and Zacks is frequency estimation, that is, assessing how often a word was presented in a list of multiple words. Frequency estimation is defined as automatic because it can be done while performing another task and is not affected by such variables as practice, intention to learn, and so forth. They argued that the ability to estimate the frequency of presentation does not decline with age, although the

This work was supported in part by Grant BNS-03711 from the National Science Foundation and by Office of Naval Research Contract No. N00014-84-K-0063, to the first author, and by a Sloan Foundation Grant to support Cognitive Science Research at Carnegie-Mellon University.

We would like to thank J. Anderson, R. Brooks, C. Dennler, L. Hasher, C. Sophian, M. Taylor, and G. Wells for commenting on previous drafts of the manuscript and C. Dennler and R. Milson for help with the analyses.

John Martin is now at Non-Procedural Systems, Westport, Connecticut.

Correspondence concerning this article should be addressed to Lynne Reder, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

ability to recall these items does. (On the other hand, Lehman and Mellinger [1984] and Kausler, Lichty, and Hakami [1984] report data that suggest that older subjects do not do as well even on "automatic tasks," that is, those tasks where conscious attention does not improve performance.)

Many investigators would not want to claim that those processes involved in text comprehension are in any way automatic. Rather, the processes that mediate recall require much cognitive effort and are, therefore, *controlled* processes. Recall performance has been shown to be affected by age. Further, if attention must be shared with another task, recall is hurt.

Mandel and Johnson suggested that many of the declines observed in memory studies may be due to focusing "too narrowly on 'traditional' laboratory stimuli and tasks." Our position is similar. We wish to show that, for identical stimuli, one can find enormous age discrepancies or none at all, depending on the task requirements or the strategies that must be used. The position we wish to examine is that those processes that are inherently easier, or perhaps more "natural," do not degrade with age.

In the first experiment, we wanted to show that older adults' ability to make verification decisions would be unimpaired if they could effectively use the *plausibility strategy* to make their judgment. By plausibility strategy, we mean using available information to infer or reason that a statement is true. But if older adults must rely on the *direct retrieval strategy*, their performance would be considerably worse than that of younger adults. By direct retrieval, we mean searching memory for a specific fact. This distinction between direct retrieval and plausibility is similar to Hasher and Griffin's (1978) distinction between *reproductive* and *reconstructive* memory retrieval.

We have examined use of plausibility versus direct retrieval with college student populations (Reder, 1979, 1982). College subjects were asked to read stories and answer questions about them. They could be asked to make recognition judgments or plausibility judgments. They could be asked the questions immediately after reading the story or at a delay. The response time and accuracy data suggest that subjects do not always use the strategy that corresponds to the task they are asked to perform. Differences in accuracy and response time to recognize presented statements that vary according to rated plausibility suggest use of the plausibility strategy. Differences in accuracy and response time to make plausibility judgments that vary according to whether the probe had been stated in the story suggest use of the direct retrieval strategy. Immediately after reading a story, college subjects were much more inclined to use the verbatim or direct retrieval strategy. As memory traces faded, subjects began to prefer the plausibility strategy. We suspected that older adults in Experiment 1 would use the plausibility strategy sooner, because their memories for specifics are typically not as good.

Experiment 1

The first experiment was designed to test the hypothesis that older subjects would rely on the plausibility strategy to a greater extent than would younger subjects, regardless of the task explicitly required of them by the experimenter. The data for the younger subjects (college students) were collected before the data for older subjects (college alumni of retirement age). The college student data are part of the data reported in Reder (1982). That experiment was conducted on a PDP-11/40 computer, whereas the older subjects were tested on an Apple II computer. The older subjects were not tested on the same computer, due to the physical location of the computer laboratory: the third floor of a building at Carnegie-Mellon which has no access by elevator. We stationed the Apple II computer in a ground-floor room that was easily accessible and more attractive in appearance.

Method

Design and procedural overview. The subjects' task was to read short stories and then answer questions about them. An example of the materials is given in Table 1. Half of the subjects in each age group were asked to make judgments concerning whether each test probe (e.g., "The heir had lost weight.") had been presented in the story. The other half of the subjects were asked to judge whether each test probe was plausible, given the information presented in the story. The material was presented on a computer display, so that we could collect response times as well as accuracy measurements.

For subjects asked to make plausibility judgments, half of the test probes were implausible in order to keep the percentage of correct positive responses at 50%. All subjects were tested on an equal number of highly plausible and moderately plausible statements. For all groups of subjects, half of the moderately and half of the highly plausible statements were explicitly stated in the stories. The other half of the moderately and highly plausible statements were implicit (not stated). Implausible statements were never stated in the story. Implausible statements were not included as test items in the recognition condition, because this would have made subjects more inclined to adopt a strategy of judging plausibility rather than to perform the prescribed task. The stated versus not-stated (explicit vs. implicit) variable gave us an estimate of the use of plausibility judg-

Table 1
Example Story and Questions Used in Experiment 1

Example story

The heir to a large hamburger chain was in trouble.
He had married a lovely young woman who had seemed to love him.
Now he worried that she had been after his money all along.
He sensed that she was not attracted to him.
Perhaps he consumed too much beer and french fries.
No, he couldn't give up the fries.
Not only were they delicious, he got them for free!
Anyway, real marital strife lay elsewhere.
His wife had never revealed before marriage that she read books.
Sometimes she used words that were many syllables long.
The proud husband decided that she was showing off.
At least, he thought, she stayed at home.
It is not too late, he resolved.
The heir decided to join Weight Watchers.
Twenty-five pounds later, the heir realized
his wife did love him after all.
He vowed never to eat another french fry.
He also told his father that he wanted no
part of his greasy food fortune.
The wife of the ex-heir smiled as they went
jogging into the sunset.
Tonight she would teach him to read.

Questions

Highly plausible
The heir got his fries from his father's hamburger chain.
The heir wanted to lose weight.
The heir had lost weight.
Moderately plausible
The heir had not worried about her motives before marriage.
The heir wished his wife did not read books.
Before marriage, his wife hid her superior intellect.

ments as a strategy. We compared false recognition of highly plausible, implicit statements with false recognitions of moderately plausible, implicit statements. Note that the different rate of false alarms as a function of plausibility cannot be due to different probabilities of making the inference during reading, and then retrieving it. The plausibility effect manifests itself even for explicit statements in both tasks, namely, faster response times and fewer errors for the highly plausible, explicit statements.

Thus far, we have described a 2 × 2 × 2 × 2 [Age: Old vs. Young × Task: Recognition vs. Plausibility × Probe Plausibility: Highly vs. Moderately Plausible (implausibles were not part of the analyses) × Stated: Explicit vs. Implicit] factorial design. Assignment of subjects to task was random within age group; whether a statement was explicitly presented in the story was randomly determined for each subject. There was yet a fifth factor in the experiment, namely, the delay between the presentation of the material to be tested and the test itself. This variable allowed us to evaluate how strategy use shifts with delay. Questions were asked either after each story (a 2–3 min delay), or after reading all 10 stories (a 20–30 min delay). This factor was manipulated between subjects.[1] Probe plausibility and the stated variable were varied within subjects. The other variables were manipulated between subjects.

Procedural specifics. Subjects read 10 stories at their own pace. They were told to read the stories in a normal fashion, as they would when reading for pleasure, and that later they would be asked some questions about the stories. The instructions also explained their assigned task. Stories were presented one sentence at a time on a computer-controlled video screen. Subjects pressed the space bar to advance to the next line of the story.

Until the subject was asked the first set of questions, there was no difference in the procedure or materials for the two tasks. In the immediate test condition, subjects judged all relevant statements for a given story, as fast and as accurately as possible, just after reading the story. For subjects in the recognition task, this meant discriminating previously presented, plausible statements from not-presented, plausible statements. After judging these statements, the next story was presented. In the plausibility task, there were twice as many items to discriminate because the subject had to separate plausible from implausible statements, ignoring whether a plausible statement had been presented as part of the story.

Subjects in the delayed test were presented with the story title prior to reading the story and prior to making timed judgments about the story. After answering the questions about a given story, another story title was displayed. The *D* and *K* keyboard keys were labeled *no* (for negative responses) and *yes* (for positive responses), respectively. Subjects were instructed to keep their index fingers on these keys at all times during the testing phase. Feedback was given regarding the accuracy of each response.

Materials. Ten stories, written by five different authors, were used. Each was approximately 20 lines long. The stories and their test questions have been used before (e.g., Reder, 1976, 1979, 1982).

The dimension of plausibility (high vs. moderate) of the plausible probes was defined previously by other subjects who rated the statements' plausibility with respect to the story. The statements were rated only by college-age students; nonetheless, the pattern of responses suggests that older subjects would rate them similarly. The statements were not explicitly stated in the story when they were rated. The implausible statements used in the plausibility judgment condition contained the same concepts used in the story, so they could not be rejected on the basis of lexical familiarity. Also, the statements were not implausible if one had not read the story.

Subjects. The young subjects were 58 Carnegie-Mellon students, 27 in the immediate condition and 31 in the delay condition. Of the 27 who were tested after each story, 14 were randomly assigned to the recognition task and 13 to the plausibility task. Of those who read all 10 stories before being tested, 14 were assigned to the recognition task and 17 to the plausibility task. Subjects received one credit toward a course requirement.

The older subjects were all Carnegie-Mellon alumni or their spouses. These alumni were contacted by mail and volunteered to come to Carnegie-Mellon to participate. The subjects' ages varied from 65 to 80 years, with a mean age of 72 years and a median age of 68. They were all in good health. Forty-six percent of the subjects had a bachelor's degree or higher in some area of science, and the other 54% had a bachelor's degree or higher in a nonscientific domain. Most subjects were retired, although some were still working at least part-time. The 49 subjects were randomly assigned to conditions. Thirteen were assigned to the immediate plausibility condition, 10 to the immediate recognition condition, 15 to the delay plausibility condition, and 11 to the delay recognition condition.

Results

We found no systematic effects for our older subjects due to either university major or employment status. Therefore, no further mention will be made of these variables. Older subjects' data were truncated at 2.5 standard deviations above the individual subjects' mean response time. Younger subjects' data were truncated at 7 s.[2] Table 2 displays the mean response times in seconds for correct responses and error rates for both the plausibility task and the recognition task at both levels of delay as a function of age. The data are broken down according to those probes that had been stated or not stated in the story and to whether they are highly or moderately plausible. Performance on the implausible statements is also given for subjects in the plausibility task. It is easier to appreciate the results by examining graphs of various portions of these data. Figure 1 displays the response times for correct decisions for the plausible statements, as a function of delay between reading and test. The data are plotted separately for both age groups and for both tasks, recognition and plausibility, but are collapsed over the plausibility of the statements and over whether the statements had been presented in the story.

A 2 × 2 × 2 × 2 × 2 analysis of variance (ANOVA) was performed on the correct response time and on the percentage of correct responses for the factors (Plausibility: Highly vs. Moderately Plausible × Presentation: Stated vs. Not-Stated × Judgment Task: Recognition vs. Plausibility × Delay: After Each Story vs. After All Stories × Age: College Students vs. Alumni Over 65). The error term used was always the interaction of subjects with the effect of interest, and is always reported for seconds rather than milliseconds. Of the 62 possible main effects and interactions across the two dependent measures, 37 were statistically significant. Rather than describing each one, later we mention only those that we consider relevant to the age contrast and our understanding of the mechanisms involved in performance differences due to age. (Consult Reder, 1982, for discussions of the results not related to age.) As expected, older subjects were significantly slower than younger subjects, $F(1, 99) = 36$,

[1] When young subjects were run in this experiment, there was a third level of delay: 2 days after reading the stories. We felt that we could not ask our older subjects to come to Carnegie-Mellon for two sessions to complete one experiment. Not only were they volunteers, they typically came to us on public transportation. Apart from the imposition, we worried about a high failure rate in returns. Therefore, we are only comparing the first two levels of delay for the younger subjects with the older subjects' data.

[2] We were reluctant to select a specific cutoff for older subjects as we had done for young subjects. Given their greater variability, the use of their own standard deviations seemed more appropriate.

$MS_e = 2.19$, $p < .01$. More interesting is the fact that the response time patterns differed for the two age groups as a function of explicit task instructions, such that the recognition task produced the slowest responses for older subjects and the fastest responses for younger subjects, $F(1, 99) = 10.4$, $MS_e = 2.19$, $p < .01$. This is consistent with our prediction that older people would find the direct retrieval strategy much more difficult to use. This pattern was complicated by a triple interaction of these factors with delay between reading and test (Age × Task × Delay), $F(1, 99) = 5.1$, $MS_e = 2.19$, $p < .01$. This triple interaction resulted from the fact that, with delay, plausibility judgments speeded up for young subjects. There is a response time slowdown for plausibility judgments with older subjects and a response time speedup for recognition judgments, but neither is significant, $t(45) = 1.112$, $SE = 1.07$ and $t(45) = .996$, $SE = 1.07$, respectively. The speedup in plausibility judgments for young subjects was significant, $t(54) = 2.0$, $SE = .26$, $p < .05$; the slowdown for recognition judgments for young subjects was not significant, $t(54) = .515$, $SE = .26$.

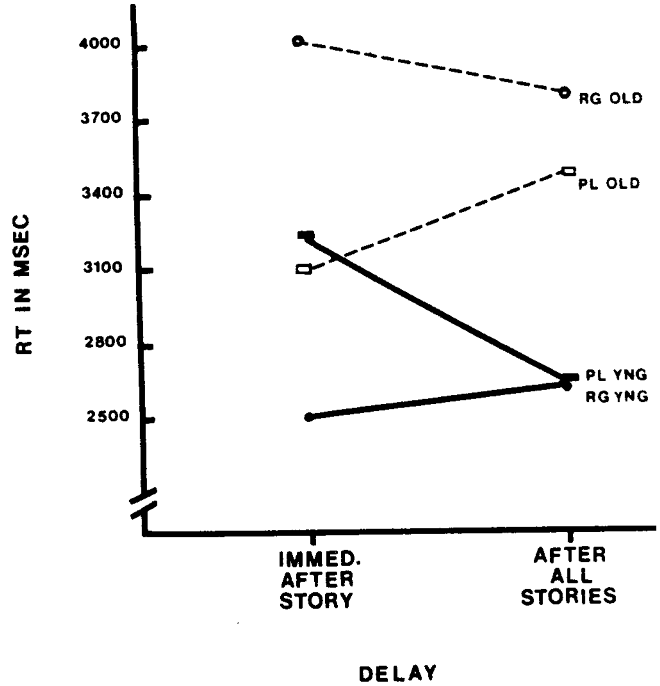

Figure 1. Experiment 1: mean reaction times (RTs) for correct responses as a function of delay between reading the story and test for both age groups in both tasks. (The data are collapsed over plausibility of the test statements and whether or not they had been presented in the story.)

Table 2
Experiment 1: Mean Response Times (RT, in Seconds) and Error Rates (ER)

	Young		Old	
Delay	Stated	Not stated	Stated	Not stated
Recognition task				
Immediate				
High plausibility				
RT	2.29	2.70	3.43	4.79
ER	.18	.21	.07	.56
Medium plausibility				
RT	2.38	2.68	3.44	4.62
ER	.14	.14	.15	.39
30-minute				
High plausibility				
RT	2.48	2.67	3.35	3.91
ER	.13	.57	.09	.70
Medium plausibility				
RT	2.66	2.77	3.35	4.44
ER	.19	.24	.15	.49
Plausibility task				
Immediate				
High plausibility				
RT	2.66	3.29	2.90	3.25
ER	.03	.08	.06	.10
Medium plausibility				
RT	2.82	4.04	2.95	3.58
ER	.08	.23	.07	.18
Implausible				
RT	3.51		3.46	
ER	.07		.18	
20-minute				
High plausibility				
RT	2.52	2.54	3.38	3.40
ER	.09	.14	.10	.11
Medium plausibility				
RT	2.58	3.08	3.25	3.94
ER	.13	.29	.12	.21
Implausible				
RT	2.79		3.72	
ER	.13		.28	

Figure 2 plots the accuracy performance for the two age groups as a function of nominal task and delay of test, again collapsing over the plausibility of the statements and whether or not the question had been presented in the story. Although older subjects were significantly less accurate overall, $F(1, 99) = 8.2$, $MS_e = .024$, $p < .01$, they were somewhat more accurate in the plausibility task than their younger counterparts. The fact that their accuracy was much worse than that of young subjects in the recognition task, but a little better in the plausibility task, produced a significant Age × Task interaction, $F(1, 99) = 14.0$, $MS_e = .024$, $p < .01$.

The response time and accuracy patterns suggest several things. First, the direct retrieval task was more difficult for older subjects than was the plausibility task. This greater difficulty was manifest in the slower response times for recognition as compared with plausibility. Second, both age groups shifted toward a greater tendency to adopt the plausibility strategy. Older subjects did not shift to more use of the plausibility strategy in the plausibility task because they already used it; however, older subjects in the recognition task did start using it more. This last point can be seen in the accuracy data displayed in Figure 3, which will be discussed later.

The interpretation of the results as a shift toward greater use of the plausibility strategy is strongly suggested by the response time data for young subjects in the plausibility task. The significant speedup in reaction time with delay for the plausibility function for young subjects was due to the not-stated items (see

Table 2). In the immediate condition, these subjects searched for the fact in memory; when the not-stated item could not be directly retrieved, they then had to judge the statement's plausibility, making this fact quite slow to verify. With delay, they adopted plausibility as a first stategy more often than they did initially. This resulted in a speedup for not-stated items in the plausibility task, because the nonproductive, direct-retrieval strategy was omitted.

Figure 3 gives further support to our theoretical interpretation. Here the accuracy data for the recognition task, for both age groups, are plotted as a function of whether or not the statement had been explicitly presented in the story, for both levels of delay. These data show that older subjects were at least as accurate as young subjects in the recognition task, just as they were in the plausibility task, so long as the test probes had been stated in the story; however, if the probes were not stated, they were appreciably worse than their younger counterparts. The Stated × Age × Task interaction was very significant for accuracy, $F(1, 99) = 29.5$, $MS_e = .024$, $p < .01$. This is because when the plausibility strategy is used in the recognition task, stated facts will be judged accurately; however, adopting the plausibility strategy in this task means that not-stated plausible statements will be erroneously accepted. For both age groups, accuracy remained relatively constant in the plausibility condition and for stated probes in the recognition task. The top two functions in Figure 3 are almost flat, whereas accuracy drops precipitously for those statements where the plausibility strategy will not work. The shift towards more frequent use of the plausibility strategy in the recognition task at longer delays caused an increase in error rates for not-stated, plausible statements, and produced a greater decline in accuracy for the highly plausible than the moderately plausible statements (see Table 2).

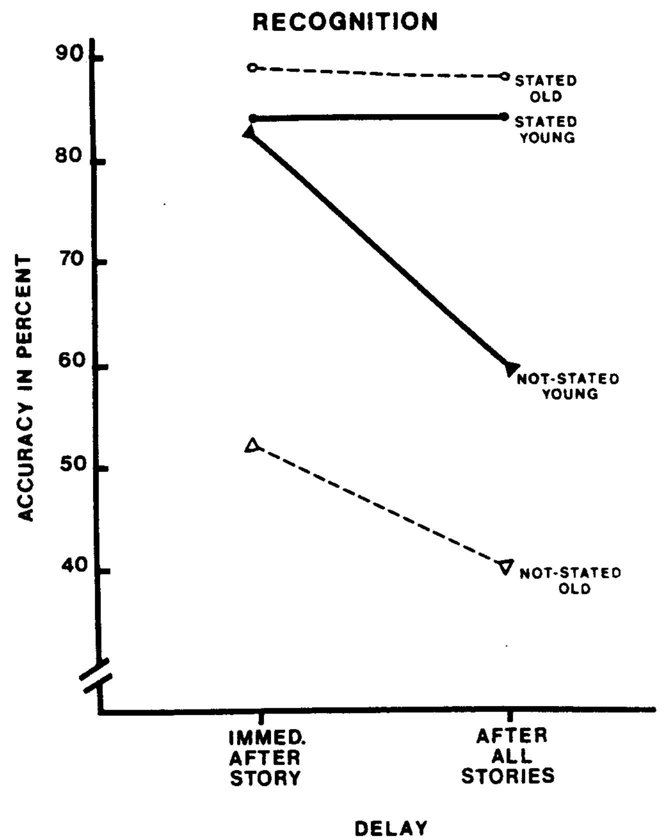

Figure 3. Experiment 1: mean percent correct for judgments in the recognition task, as a function of whether or not the statement had been presented in the story, delay of test, and the age group of the subjects. (The data are averaged over the plausibility of the test statements.)

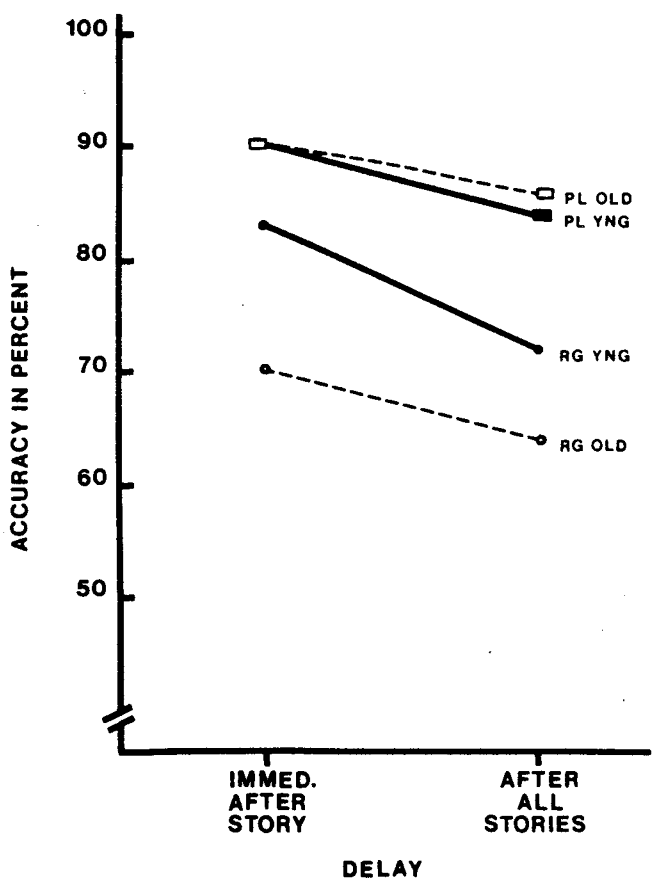

Figure 2. Experiment 1: mean percent correct for judgments at both levels of delay for both age groups in both tasks. (The data are averaged over the plausibility of the test statements and whether or not they had been presented in the story.)

Discussion

These data support the hypothesis that older adults can effectively use the plausibility strategy but perform poorly when the direct retrieval strategy is required. This avoidance or inability to use the direct retrieval strategy causes older subjects to err more often in the conditions where the probe has not been stated and the official task is to make recognition judgments.

There are two explanations for why older subjects rely on the plausibility strategy more than the direct retrieval strategy. One explanation, given earlier, is that searching for a specific fact is harder for older adults because their memory traces fade faster than those of younger adults. At short delays, the direct retrieval strategy is easy for young subjects because verbatim traces are likely to be readily available for them. This interpretation is consistent with a result of Cohen and Faulkner (1981) which showed that older subjects lose surface structure information more quickly than do younger adult subjects.

Another possibility that is also consistent with the results is that older subjects simply find the direct retrieval strategy very difficult to execute. Their greater tendency to use the plausibility strategy could be unrelated to whether or not surface information is still available. Possibly, older subjects prefer a plausibility-like strategy because it is a more "natural" strategy, a practiced strategy, and therefore, an easier, better performed strategy. The next experiment was designed to discriminate between these two explanations by using a task where age differences in encoding or retention of surface structure were irrelevant. The experiment used a "semantic memory" task rather than an "episodic memory" (see Tulving, 1972) task. If differences in strategy use are due to weaker encoding and retention of new facts, then the same strategy-use differences would not necessarily occur. If the age variable does not show the same pattern, then the strategy-effects are probably due to weaker encoding or memory traces. Our hypothesis was that performance differences between age groups would be minimal when the task did not require a careful inspection of memory traces. We suspected that when careful inspection of memory associations was required, large age differences would become apparent.

Experiment 2

To test the hypothesis that the different pattern of data across age is not due to differences in the strength of episodic memory traces, but rather to *ease of strategy use*, the second experiment used a semantic memory task. The test stimuli involved common category names and instances of the categories. Presumably, this information was learned by everyone at roughly the same age. Although one might argue that older subjects' traces have faded over time, one could also argue that they have had more opportunities to practice this information, leading to stronger traces. Indeed, there is evidence that vocabulary is relatively impervious to other deleterious effects of age (e.g., Gardner & Monge, 1977; Schaie, 1958; Thorndike & Gallup, 1944).

Subjects were required to make simple judgments in this experiment. Two words were displayed on the screen simultaneously, and the top word was always a category name. The subject's task was to decide if the bottom word was an instance of that category or not. When the bottom word was an instance of the category, half of the time it was a dominant instance of the category (i.e., one which is frequently associated with the category) and the other half of the time it was a less frequent, less dominant instance of the category. For non-instances, half of the time the non-instance was highly related to the category name, for example, *disease-medical* or *furniture-house*, and the other half of the time it was unrelated to the category name, for example, *disease-gravel* or *furniture-alphabet*. If older people have a greater tendency to adopt a plausibility-like strategy, then they would tend to respond positively to related non-instances as well as to positive instances of the category. If the bottom term shares many features with the category name, it is *plausible* that the term is an instance of it. It is worth noting that this experiment bears some resemblance to one reported by Eysenck (1975), except that he did not vary the relatedness of the distractors. That manipulation, of course, is critical to testing our hypothesis.

Table 3
Example of Word Pairs Used in Experiment 2

Associated strength	Example
Instance (yes response)	
Dominance	
High	Country Russia
Low	Country Iceland
Noninstance (no response)	
Relatedness	
High	Country Continent
Low	Country Wrench

Method

Subjects. Alumni of the Carnegie Institute of Technology and Carnegie-Mellon University and their spouses were used as subjects. The older subjects varied in age between 64 and 75 years, with a mean age of 70. The young subjects varied between 20 and 31 years, with a mean age of 27. This experiment used 8 young subjects and 10 older subjects. Alumni, particularly young alumni, who were willing to participate were hard to come by. Therefore, we made do with small numbers of subjects. All subjects were contacted by mail and volunteered to participate in our experiment, that is, we used no current college students. Subjects in both the young and old age groups had majored in either engineering, science, or fine arts and had careers in areas matching their majors. All subjects held bachelor's degrees and some held master's degrees.

Materials. Fifty-four categories were chosen from the Battig and Montague (1969) norms. (In some cases, a shorter form of the category name was substituted so that it could be expressed in one word. For example, the category name *type of clothing* was changed to *clothing*.) The high-dominance and low-dominance instances were also derived from the Battig and Montague norms. High-dominance instances were typically the words with the highest or second highest frequency rating. For example, the high-dominance instance for BIRD was ROBIN. Low-dominance instances were selected from the words with the lowest frequency ratings. For BIRD, the word STORK was used. Table 3 gives an example of each condition used in Experiment 2.

For the non-instances, we had to develop our own norms. We asked 50 undergraduate students to complete questionnaires which listed the 54 category names. Beside each name were four blanks. Students were requested to write in the first four associations that came to mind. From these 50 questionnaires, we selected the noun most frequently mentioned for each word that was not an instance of the category. For BIRD, FLY was the highly related non-instance. In selecting nonrelated, non-instances for each category name, we tried to choose a word that matched the related non-instance in length and frequency, yet seemed to us to be totally unrelated to the category name. This item could not have been generated by any subject completing the questionnaire, nor be listed in the Battig and Montague norms. For example, BOX was used as an unrelated non-instance of BIRD.

The Battig and Montague norms were constructed from ratings generated by college students. Therefore, we had some concern about the validity of these indices for our older subjects. Howard (1979), however, has shown that the Battig and Montague norms are representative for older subjects as well as younger subjects. She asked various age groups to perform the task that Battig and Montague had originally asked only of college students, and found considerable stability across age groups.

Procedure and design. For each subject, the computer randomly selected 48 category names from the possible 54. Twelve of these category names were assigned to each of the four conditions: high-dominance instance, low-dominance instance, highly related non-instance, and not-related non-instance. Subjects from both age groups were run individually on an Apple II Plus computer. They were instructed to decide if the bottom word on the screen was an instance of the word on the top. Subjects initiated each trial by pressing the middle button of a three-button box, which caused the word pair to appear on the screen and a clock to start. The outer two buttons were labeled *yes* and *no*, which subjects used to indicate their responses. They were instructed to rest their index fingers on the response buttons and to respond as "quickly as possible without sacrificing accuracy." The response and the time to make the decision were stored by the computer for each trial, and subjects were given feedback about the accuracy of their responses. Prior to beginning the critical trials, five simple, unrelated questions were asked of the subjects in order to familiarize them with the response buttons.

Results and Discussion

Figures 4 and 5 graphically display the data for Experiment 2, broken down as a function of age group and associative strength of the bottom term to the category name. Figure 4 plots the accuracy data in percentage. The left-hand panel is for instances, whereas the right-hand panel is for non-instances. Figure 5 plots

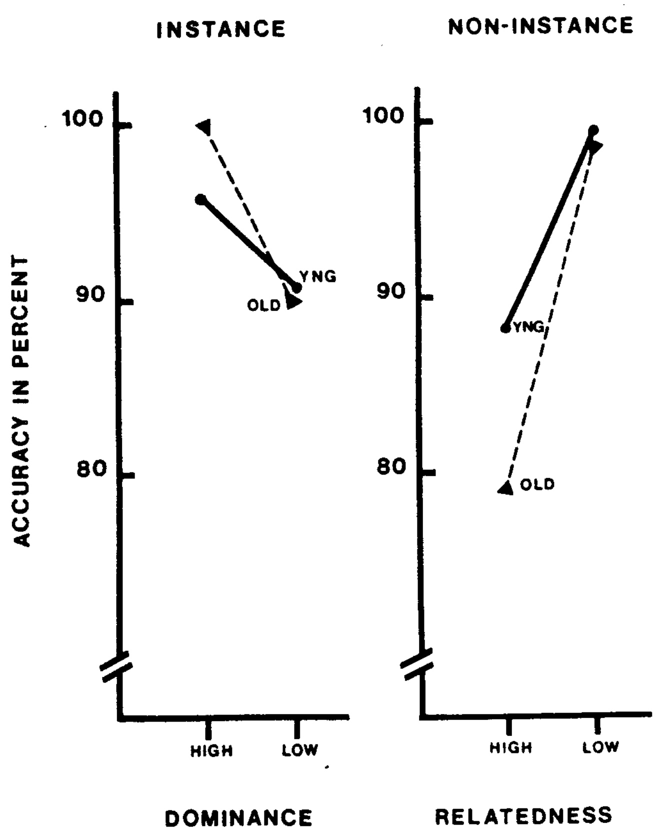

Figure 4. Experiment 2: Mean percent correct as a function of age and dominance or relatedness.

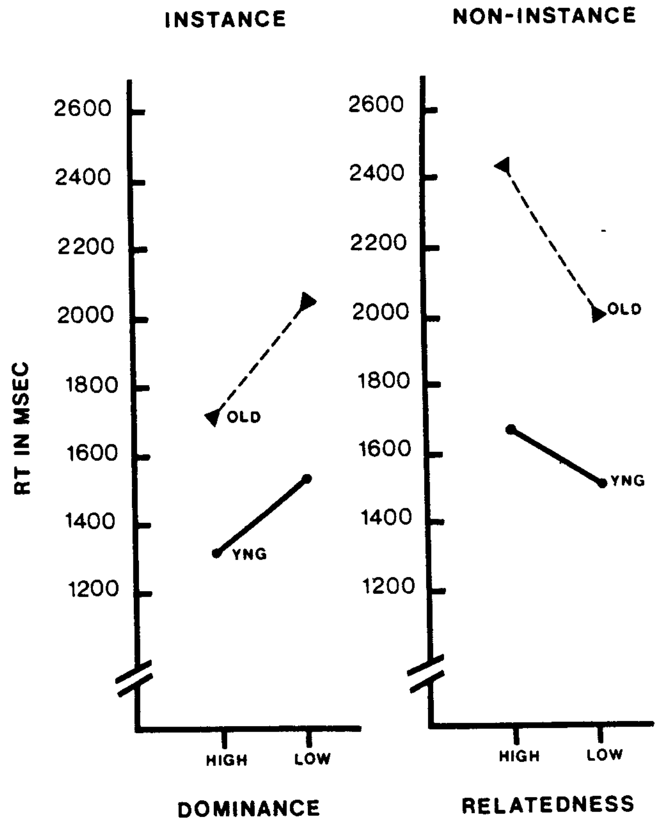

Figure 5. Experiment 2: Mean reaction time (RT) for correct responses as a function of age and dominance or relatedness.

the reaction time data in milliseconds in the same manner. Mean reaction times were calculated only for correct responses. Response times were truncated to 2.5 standard deviations above each subject's overall mean. ANOVAs were performed on both the accuracy data and the response time data, using the factors of age, response type (instance vs. non-instance), and associative relatedness (*high* means dominant for instances and highly related for non-instances). The error term used was always the interaction of the subjects with the relevant factor. The error terms for response times are reported in seconds.

Consider the accuracy data first. There was a clear effect of associative relatedness, such that subjects were more accurate for high-dominant than low-dominant items in the instance condition, and more accurate for unrelated than related items in the non-instance condition. Because the pattern of associative relatedness is reversed depending on type of correct response (instance vs. non-instance), this effect was seen as a significant interaction of relatedness or dominance with type of correct response, $F(1, 16) = 28.8$, $MS_e = 1.12$, $p < .01$. There also appeared to be an Age × Type of Response interaction: Older subjects were more accurate than younger subjects for instance judgments, but less accurate for non-instances. This interaction, however, was not significant, $F(1, 16) = 2.6$, $MS_e = 1.12$.

What is clear from these data is the following: Older subjects' accuracy is essentially indistinguishable from younger subjects' accuracy, so long as a plausibility-like judgment process will produce the correct response. When degree of relatedness can be used to correctly infer that an item is or is not an instance of the category, then older subjects' accuracy is as good as or better than young subjects'. On the other hand, when the relatedness of the item predicts the wrong response, all subjects perform badly, but older subjects perform much worse than young subjects. Although no main effect or interaction involving age was significant for accuracy, the difference between young and old for related non-instances was significant, $t(16) = 2.26$, $SE = .50$, $p < .05$. One might view the data in the following way: In all conditions except the highly related, non-instance (*country-continent*) condition, a lenient criterion for acceptance produces the right answer; therefore, older subjects are at least as accurate as younger subjects. However, when the items are "tricky," older subjects are more easily "tricked." The strategies people maintain throughout their life-span are those needed in situations that occur in everyday life, rather than in contrived experimental tasks. The highly related non-instances are not seen in everyday life.

The response time data present a similar pattern and the statistics are more powerful. There was, of course, a main effect of age, such that older subjects were slower than young subjects, $F(1, 16) = 10.6$, $MS_e = .5184$, $p < .01$. We find this unsurprising even though older subjects responded faster than younger subjects in some conditions in Experiment 1. The reason older subjects could sometimes respond faster than younger subjects in that experiment was due to the greater tendency of young subjects to use an inappropriate first strategy before executing the appropriate one. Had these subjects not gone on to try the appropriate second strategy, younger subjects would have had faster response times in all conditions.

We were also not surprised that subjects also took longer to reject non-instances than instances, $F(1, 16) = 30.5$, $MS_e = .0303$, $p < .01$. Subjects were, in general, faster to accept high- than low-dominance instances, and faster to reject low- than high-relatedness non-instances, producing an interaction of associative relatedness with response type (instance vs. non-instances), $F(1, 16) = 74.6$, $MS_e = .0203$, $p < .01$.

Of special interest is the triple Age × Response Type × Associative Relatedness interaction, $F(1, 16) = 7.2$, $MS_e = .0203$, $p < .05$. Analogous to the accuracy data, older subjects performed appreciably worse on the highly related non-instances.

In summary, the results of this experiment indicate that when relatedness effects can facilitate judgment, older subjects' accuracy is at least as good as younger subjects': When it gives the wrong response, older subjects' performance is worse. Relatedness seems to be used as a heuristic for making a kind of plausibility judgment: When the items are highly related, it is plausible that the first item is an instance of the category named by the second. This heuristic, using relatedness to make plausible inferences about category membership, is used to some extent by the young subjects as well as the older subjects.

General Discussion

A consistent pattern has emerged from these two experiments. In both studies, older subjects perform as accurately as young subjects when plausible reasoning can facilitate a response; when the relatedness-heuristic or other plausibility process produces the wrong response, older subjects' performance is worse. In Experiment 1, older subjects were faster than younger subjects in situations where they were more inclined to adopt the plausibility strategy, and the careful search of memory strategy was ineffective. Their accuracy was always as good as or better than that of their younger counterparts, provided that the plausibility strategy produced the correct response. In Experiment 2, older subjects were appreciably slower and less accurate when the relatedness judgment produced the wrong response. In other conditions, older subjects' accuracy again rivaled or bettered young subjects'.

The good performance of older subjects relative to younger subjects can be explained if we make two assumptions. The first is that the plausibility strategy is less demanding of attention than the direct-retrieval or careful-inspection strategy. The second is that processes that are not attention demanding do not degrade with age.

The first assumption, that plausibility judgments impose a less demanding processing load than direct retrieval, may not seem intuitive. One might think that making plausibility judgments requires complex inferential processes and, therefore, is more demanding. However, we are not referring to "deductive" or "logical" reasoning in this context. Rather, we imagine two types of processes for making plausibility judgments. One is a *partial match strategy*, where, if the "overlap" is high between features in the query and in the memory structure, the query is seen as plausible or consistent with memory. Consider the question, "How many animals of each kind did Moses take on the Ark?" Most subjects answer "two" (Erikson & Mattson, 1981), although they know that Noah, not Moses, built the ark. The reason most people err when asked this kind of "trick" question is that our normal mode of processing is a partial match strategy. If there is a high overlap of relevant features, people assume a match. It is in the exceptional situation that people feel compelled to closely examine propositions and carefully match the structures with those in memory. We assert that partial matching is automatic, whereas careful matching requires controlled operations. The idea that plausibility judgments require less processing is reminiscent of a processing resource account of age differences in recall (Rabinowitz, Craik, & Ackerman, 1982) that suggests older people are more inclined to use a general or stereotyped encoding of information.

The second type of plausibility mechanism that we view as nondemanding or automatic involves matching queries to plausibility rules, where certain conditions have to be met for a statement to be considered plausible. Consider being asked to answer the question, "Was Dorothy a nice girl in *The Wizard of Oz*?" We might have plausibility rules stored of the forms, "A person who wants to help others is nice," and "A child who loves his/her caretakers (Auntie Em) is nice." A lot of information from the story would allow those rules to apply.

Tasks requiring controlled attention produce the greatest differences between younger and older subjects, and also provide the most room for individual differences. This would imply that tasks that required an exact match (the recognition task in Experiment 1) or careful inspection of a relationship (the related non-instances in Experiment 2) would be more likely to degrade with increasing age than the plausibility judgments.

Cohen and Faulkner (1983) looked at age differences in mental rotation and picture-sentence verification tasks. They also found that "the age difference was minimized when older adults adopted strategies that reduced the amount of processing." They did not observe differences in choice of strategies as a function of age; however, they felt that "the number of participants was too small to license any general claims about the absence of age differences in strategy selection" (p. 454).

The conclusion that older subjects' performance is especially bad for tasks that are especially difficult is well established and not controversial. We mentioned in the introduction that recently a number of researchers have made the distinction between automatic processes that do not degrade with age, and effortful processes that do. One reason that certain processes might demand little attention and be automatic is due to extensive practice. It certainly follows that people are more prone to practice those processes that are easier, which might accentuate differences with age. Salthouse (1982) has argued that age differences may arise due to disuse of a particular procedure or process. We would argue that monitoring for "catch" questions (i.e., related non-instances in Experiment 2) is probably not practiced, except in school. Similarly, we do not have to practice discriminating plausible, implicit statements from plausible, explicit statements in everyday situations. What we do practice is making plausibility or verification judgments, namely, "Does this follow from what I read?"

The interpretation that we impose on our results may help to explain why other studies looking at inferential processes as a function of age are inconsistent with each other. Belmore (1981) found that older subjects were equally good at verifying inferences and paraphrases of short prose passages and did not differ from younger adults. Cohen (1979) found the opposite result, that older subjects were worse at making inferences than at recalling asserted information. Both experimenters presented subjects with very short passages, although Belmore asked for verification and Cohen asked for short answers. The critical difference was the kind of information that had to be held in memory in order to make the inferential judgment. Belmore's inferences involved reasoning using world knowledge. Cohen's passages required deductions based entirely on the premises presented in the passage. For example, after reading sentences that described the layout of the house and then mentioned which people were in which rooms, Cohen's subjects might have to say who was disturbed by traffic. Nothing in the material would motivate the reader to combine those premises (sentences) prior to being asked the question, and the conclusion would not be inferrable from world knowledge. In Belmore's passages, one of the premises would be implicit knowledge about how the world works, for example, if banks are already closed for the weekend, today must be Friday. Belmore's inferences, like ours, involved relating story information to world knowledge to verify a statement. Our position is that older subjects are good at making inferential or plausibility judgments so long as they are not required to retrieve specific details. Cohen's task was particularly demanding of just that sort of specific fact retrieval.

Finally, it should be pointed out that there are two somewhat different theoretical accounts of the results reported here that are not easily distinguished. One is that older subjects do poorly in certain contexts because they are unable to make careful matches to memory. The other explanation is that older subjects do poorly in certain situations because they are unwilling to use the direct retrieval or careful inspection strategy, but could use it if they became motivated. We have no evidence that older subjects could use the direct retrieval strategy effectively if they wanted to. On the other hand, past research strongly suggests that young subjects initially have a strong tendency to use direct retrieval, but then switch to a plausibility mechanism when the queried information has not been seen recently (Reder, 1982). In other words, older subjects just have a stronger inclination to use the strategy that younger subjects also tend to adopt.

The notion that age differences in memory performance might be attributable to differences in strategy use is not a unique hypothesis. For example, Perlmutter and Mitchell (1982) claim that older subjects' major problems are due to "inefficient spontaneous use of encoding and retrieval strategies" (p. 143), and are not due to deficits in their encoding abilities, per se. However, others have claimed that it is older adults' effective use of strategies that allows them to compensate for inferior memory abilities (e.g., Moscovitch, 1982). Moscovitch found that older adults are more likely to write down appointments so as not to forget them and, consequently, miss fewer. Indeed, Rabbitt (1982) suggests that the major task in aging research is to describe the ways in which people actively optimize their performance to cope with changing task demands and to circumvent or minimize growing failures in their own efficiency.

We do not completely understand the relative contributions of memory loss versus an inability or unwillingness to use a careful inspection strategy on performance decrements. Nonetheless, we have illustrated situations for which older subjects can perform as well as college-aged subjects. The typical or everyday tasks that allow people to use automatic, partial-match processes, as opposed to those that require careful, controlled-matching processes, seem unaffected by age.

References

Attig, M., & Hasher, L. (1980). The processing of frequency of occurrence information by adults. *Journal of Gerontology, 35*, 66–69.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology Monograph, 80*, 1–46.

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

Birren, J. E., & Schaie, K. W. (Eds). (1977). *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold.

Botwinick, J., & Storandt, M. (1974). *Memory, related functions, and age*. Springfield, IL: Charles C Thomas.

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

Cohen, G., & Faulkner, D. (1981). Memory for discourse in old age. *Discourse Processes, 4*, 253–265.

Cohen, G., & Faulkner, D. (1983). Age differences in performance on two information-processing tasks: Strategy selection and processing efficiency. *Journal of Gerontology, 38*, 447–454.

Craik, F. I. M. (1968). Types of error in free recall. *Psychonomic Science, 10*, 353–354.

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

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. Craik & S. Trehaub (Eds.), *Aging and cognitive processes: Vol. 8. Advances in the study of communication and affect* (pp. 191–210). New York: Plenum Press.

Craik, F. I. M., & Masani, P. A. (1967). Age differences in the temporal integration of language. *British Journal of Psychology, 58,* 291–299.

Erber, J. T. (1974). Age differences in recognition memory. *Journal of Gerontology, 29,* 177–181.

Erikson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20,* 540–552.

Eysenck, M. W. (1975). Retrieval from semantic memory as a function of age. *Journal of Gerontology, 30,* 174–180.

Gardner, E. F., & Monge, R. H. (1977). Adult age differences in cognitive abilities and educational background. *Experimental Aging Research, 3,* 337–383.

Hasher, L., & Griffin, M. (1978). Reconstructive and reproductive processes in memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 318–330.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

Howard, D. (1979). *Category norms for adults between the ages of 20 and 80* (Tech. Rep. NIA-79-1). Washington, DC: Georgetown University, Cognition Laboratory.

Hunt, E., & Hertzog, C. (1981). *Age related changes in cognition during the working years* (Final Report for ONR Contract No. N00014-79-C-0345). Seattle: University of Washington, Department of Psychology.

Kausler, D. H. (1982). *Experimental psychology and human aging.* New York: Wiley.

Kausler, D. H., & Hakami, M. K. (1982). Frequency judgments by young and elderly adults for relevant stimuli with simultaneously present irrelevant stimuli. *Journal of Gerontology, 37,* 438–442.

Kausler, D. H., Lichty, W., & Hakami, M. K. (1984). Frequency judgments for distractor items in a short-term memory task: Instructional variation and adult age differences. *Journal of Verbal Learning and Verbal Behavior, 23,* 660–668.

Laurence, M. W. (1967). A developmental look at the usefulness of list categorization as an aid to free recall. *Canadian Journal of Psychology, 21,* 153–165.

Lehman, E. B., & Mellinger, J. C. (1984). Effects of aging on memory for presentation modality. *Developmental Psychology, 20,* 1210–1217.

Mandel, R. G., & Johnson, N. S. (1984). A developmental analysis of story recall and comprehension in adulthood. *Journal of Verbal Learning and Verbal Behavior, 23,* 643–659.

McCormack, P. D. (1981). Temporal coding by young and elderly adults: A test of the Hasher-Zacks model. *Developmental Psychology, 17,* 509–515.

Moscovitch, M. (1982). A neuropsychological approach to perception and memory in normal and pathological aging. In F. Craik & S. Trehaub (Eds.), *Aging and cognitive processes: Vol. 8. Advances in the study of communication and affect* (pp. 55–78). New York: Plenum Press.

Perlmutter, M., & Mitchell, D. B. (1982). The appearance and disappearance of age differences in adult memory. In F. Craik & S. Trehaub (Eds.), *Aging and cognitive processes: Vol. 8. Advances in the study of communication and affect* (pp. 127–144). New York: Plenum Press.

Poon, L. W., Fozard, J. L., Cermak, L. S., Arenberg, D., & Thompson, L. W. (Eds.). (1980). *New directions in memory and aging: Proceedings of George Talland Memorial Conference.* Hillsdale, NJ: Erlbaum.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium.* Hillsdale, NJ: Erlbaum.

Rabbitt, P. M. A. (1982). How do old people know what to do next? In F. Craik & S. Trehaub (Eds.), *Aging and cognitive processes: Vol. 8. Advances in the study of communication and affect* (pp. 79–98). New York: Plenum Press.

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

Reder, L. M. (1976). *The role of elaborations in the processing of prose.* Unpublished doctoral dissertation, University of Michigan.

Reder, L. M. (1979). The role of elaborations in memory for prose. *Cognitive Psychology, 11,* 221–234.

Reder, L. M. (1982). Plausibility judgments vs. fact retrieval: Alternative strategies for sentence verification. *Psychology Review, 89,* 250–280.

Salthouse, T. A. (1982). *Adult cognition: An experimental psychology of human aging.* New York: Springer-Verlag.

Schaie, K. W. (1958). Rigidity-flexibility and intelligence: A cross-sectional study of the adult life span from 20 to 70. *Psychology Monographs, 72,* (462, Whole No. 9).

Schonfield, D., & Robertson, B. A. (1966). Memory storage and aging. *Canadian Journal of Psychology, 20,* 228–236.

Thorndike, E. L., & Gallup, G. H. (1944). Verbal intelligence of the American adult. *Journal of General Psychology, 30,* 75–85.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press.

Walsh, D. A., & Baldwin, M. (1977). Age differences in integrated semantic memory. *Developmental Psychology, 13,* 509–514.

Walsh, D. A., Baldwin, M., & Finkle, T. J. (1980). Age differences in integrated semantic memory for abstract sentences. *Experimental Aging Research, 6,* 421–444.

Zacks, R. T. (1982). Encoding strategies used by young and elderly adults in a keeping track task. *Journal of Gerontology, 37,* 203–211.

Received February 18, 1985
Revision received June 3, 1985 ■