Psychology and Aging
1993, Vol. 8, No. 2, 176–186


0882-7974/93/$3.00

# Individual Differences in Cross-Sectional and 3-Year Longitudinal Memory Performance Across the Adult Life Span

Elizabeth M. Zelinski, Michael J. Gilewski, and K. Warner Schaie

This article reports individual differences analyses of performance on list and prose memory tasks for 250 men and 258 women aged 55–84. Being retested, higher reasoning and vocabulary scores, and female gender predicted better prose recall and list recognition performance. For list recall, retest status, age, years of schooling, and gender, as well as reasoning and vocabulary, were reliable independent predictors. After 3 years, 106 men and 121 women returned for a retest. Analysis of individual differences in 3-year performance indicated that, once Time 1 performance had been partialed, individual change could be predicted by age or reasoning, but neither variable uniquely accounted for change. Analysis of data of individuals who experienced considerable decline or improvement in 3-year scores indicated that decline was consistently associated with advanced age. Ramifications for theoretical models in memory research are discussed.

Implicit in the methodology of many aging studies is the assumption that group means in cross-sectional studies reflect changes in performance indicative of age-related declines in episodic recall. The search for theoretical mechanisms to explain these declines, however, has too often resulted without providing additional insight into their source, because, in many cases, age does not consistently interact with the manipulations in which the purported mechanism can be observed (Light, 1991). It has been suggested that an individual differences approach, in which cognitive factors believed to underlie performance are examined correlationally, may yield more informative results than experimental approaches relying on group differences (e.g., Dwyan & Jacoby, 1990; Hultsch & Dixon, 1990).

The correlational approach has long been supported by developmental research methodologists, who have suggested that individual differences in performance by older adults and the mechanisms underlying such differences have been obscured by group data analysis (e.g., Baltes & Nesselroade, 1979). There is also a growing interest in uncovering such differences, as seen in recent articles (e.g., Cockburn & Smith, 1991; Hartley, 1986; Hultsch, Hertzog, & Dixon, 1990; Rabbitt, 1989; Rice & Meyer, 1986; Zelinski, Gilewski, & Anthony-Bergstone, 1990).

## Individual Differences Approaches to Memory

Studies of individual differences in memory point to several interrelated cognitive factors that predict performance. These include basic processing variables, such as verbal speed (Hunt, 1978), as well as products of cognition, such as reasoning ability (Kyllonen & Christal, 1990; Woltz, 1988). It has also long been assumed that breadth of knowledge is predicted on level of fluid abilities (Horn, 1968), because knowledge and reasoning levels are intercorrelated (Kyllonen & Christal, 1990; Larson & Saccuzzo, 1989).

Theorists have made some attempts to determine which mechanisms underlie specific abilities. For example, Salthouse (1991) has reviewed the role of memory processes in reasoning in old adults and concluded that memory as well as abstraction deficits may explain age differences in reasoning. However, high intercorrelations of memory and reasoning make it difficult to isolate which of these factors is more cognitively "primitive" (i.e., subsumes the other). A parallel problem is observed in the study of the mechanisms of verbal ability. Because of this, Sternberg (1985) has suggested that verbal intelligence levels are the result of interactions between basic elementary processes and complex products of cognition, whereby these function essentially as bootstrapping operations, building on each other.

Similarly, in investigating the role of ability in memory processes, we take the position that knowledge and reasoning are both products of memory and other operations and, once developmentally established, play an important role in memory themselves. For example, the use of mnemonic strategies boosts recall, and abstract problem-solving ability is used to identify and invoke a mnemonic strategy by the rememberer. Thus, an individual with high reasoning performance would be expected to have high episodic memory performance, and vice versa.

In the aging literature, fluid intelligence abilities, as measured by reasoning tasks, are good predictors of memory performance (Cockburn & Smith, 1991; Rabbitt, 1989). There is also evidence that working memory capacity, which is correlated in the .90s with reasoning (Kyllonen & Christal, 1990; Larson & Saccuzzo, 1989), reliably predicts memory in older populations (e.g., Hultsch et al., 1990).

Elizabeth M. Zelinski, School of Gerontology, University of Southern California; Michael J. Gilewski, Cedars-Sinai Medical Center, Los Angeles, and University of Southern California; K. Warner Schaie, Department of Human Development and Family Studies, Pennsylvania State University.

This research was supported in part by Program Grant AG0133 to the Andrus Gerontology Center, University of Southern California, and by Grant R01 AG4114 to Elizabeth M. Zelinski, both funded by the National Institute on Aging.

Correspondence concerning this article should be addressed to Elizabeth M. Zelinski, School of Gerontology, University of Southern California, Los Angeles, California 90089-0191.

On the other hand, measures of crystallized abilities, such as vocabulary knowledge, are somewhat less successful predictors. Vocabulary scores are typically used to match differing age groups in achievement, but many experiments report age differences in favor of the young on episodic memory measures. Studies directly examining vocabulary as a predictor of memory performance relative to fluid abilities show either no reliable relationships (Cockburn & Smith, 1991; Hultsch et al., 1990; Rabbitt, 1989) or small amounts of variance accounted for (Hartley, 1986).

## Age Versus Ability in Memory Research

The question of the extent to which the process of aging accounts for variance in memory performance over and above that of fluid or crystallized abilities has major implications for an understanding of the mechanisms of memory performance in older adults. In young adults, those high in fluid abilities such as reasoning tend to perform at high levels in memory tasks. One might expect this to be true of older adults as well. But is there some process inherent in aging, not yet identified, that also affects memory? If so, one would expect to find that once the effects of reasoning and vocabulary and other known predictors of memory are partialed out, "age" should still reliably account for variance in memory. One would then need to identify changes with age, once other predictors have been partialed, that relate to memory.

Such findings would argue that it is necessary to identify sources of change related specifically to the aging process as well as sources of variance in abilities in adults of all ages. For example, the model of memory deficits in conscious recollection proposed by Light (e.g., Light, 1991) argues for age-related changes independent of initial level of performance; however, findings from individual differences research that abilities such as reasoning and vocabulary are important predictors of performance would also require that the effects of those abilities be accounted for before examining whether mechanisms such as conscious recollection are the major primitive source of age differences.

If the age variable does not account for additional variance in memory once abilities have been partialed, it would suggest that these fluid or crystallized abilities are the source of age-related memory decline and that the factors underlying age-related changes in these abilities would need to be elucidated. For example, Stankov (1988) has argued that attention underlies many cognitive functions and that declines in attention could account for many of the effects of cognitive aging. If individual differences in specific intellectual abilities are more important than the age variable in predicting memory performance, cognitive aging psychologists would want to focus on sources of psychometric performance change in older adults to understand memory decline.

Finally, if age interacts with fluid or crystallized ability, differential decline approaches to studying memory would be called for, whereby mechanisms underlying both age- and ability-related changes would have to be investigated using interactive models of individual differences, such as those proposed by Aiken and West (1991). Here, performance at different levels of age and ability would need to be assessed independently to understand memory functioning, and the multiplicative effects of various mechanisms of memory could be identified.

In summary, the use of individual differences models examining the relative contributions of age- and ability-related mechanisms in memory research would probably reflect the complexity of the processes underlying memory performance to a much greater extent than the experimental designs currently in wide use. Certainly they would complement current research approaches and help to identify new questions in the search for mechanisms of memory change that occur with age.

In examining the empirical data on the role of chronological age and fluid or crystallized abilities as predictors of memory performance, one finds mixed results. Some studies conclude that age accounts for little, if any, variance, in contrast to the amount accounted for by both fluid abilities and products of cognition (Hultsch et al., 1990; Rabbitt, 1989). Others indicate that, independent of abilities, age accounts for 10%–14% of the variance (Cockburn & Smith, 1991; Hartley, 1986).

One of the reasons for discrepancies in findings is that the studies in which the effects of age were minimized had large sample sizes and relatively small ratios of the number of predictors to subjects (Hultsch et al., 1990; Rabbitt, 1989). On the other hand, those in which age was a significant predictor had fewer than 100 subjects each and had relatively high predictor–subject ratios (Cockburn & Smith, 1991; Hartley, 1986). Because high predictor–subject ratios may result in unstable regression equations, the results of the latter studies may not be as generalizable and reliable as might be assumed, despite their statistical significance (Cohen & Cohen, 1975).

Another problem inherent in examining the question of age versus abilities as unique sources of variance in performance is that there may be methodological difficulties in determining which of a pair of highly correlated variables, such as age and reasoning, is the true underlying explanatory construct.

In the study by Hultsch et al. (1990), the method of testing the relative contributions of variables to $R^2$ was based on a set of models comparing the variance accounted for by age when it was entered first as a predictor of memory with that accounted for when it was entered subsequent to a series of other predictors of memory performance. In the latter model, variance in performance due to the previously entered predictors had already been partialed out, and the reliability of the prediction of age for the residual variance was then evaluated. As age was entered later into the equation, it accounted for relatively little variance, providing evidence for the conclusion that individual differences in the partialed abilities are more important than age in determining memory function.

Although such an approach is methodologically sound, there are no direct comparisons of the variance accounted for when predictors that are highly intercorrelated (e.g., age and reasoning) are tested early and late in the regressions. This is a more direct form of hypothesis testing in which potential explanatory variables "compete" as predictors (see also Cohen & Cohen, 1975) and the statistically reliable amount of variance accounted for serves as an operationalization of the relative importance of predictors.

In our study, we extended previous work by examining an array of individual differences variables as predictors of memory, including those of reasoning, spatial ability, and vocabulary. We compared two hierarchical regression models, one with age entered early and one with reasoning, space, and vocabulary abilities entered early, to examine the reliability of the unique variance associated with each category of predictor.

## Longitudinal Research on Memory and Aging

The second issue that we addressed in our data was whether there are individual differences in patterns of 3-year change. Few studies have investigated this question, which is important clinically because it is assumed that memory declines in individuals may be due to "benign senescent forgetting" (Kral, 1962) or "age-associated memory impairment" (Crook et al., 1986), signs of normal declines in memory, or to the early manifestation of dementing diseases. When longitudinal change is examined clinically, error in measurement is rarely assumed to be the cause of declines, although, without multiple assessments, instability in performance cannot be ruled out.

Salthouse, Kausler, and Saults (1986) questioned the likelihood of obtaining findings of individual longitudinal declines in performance in older adults because the amount of within-subjects variance in test performance in a single session far exceeds estimates for between-subjects differences that could be attributable to aging, even for tests as reliable as the Wechsler Adult Intelligence Scale (Wechsler, 1958) Digit Span. It may therefore be difficult to observe statistically significant changes in performance at the level of individual subjects.

Regarding patterns of longitudinal memory change, little is known, and data have been analyzed only at the group level. In the Baltimore Longitudinal Study of Aging (BLSA), subjects were evaluated at points approximately 6 years apart on paired associates and serial learning, as well as visual memory (see Arenberg, 1983, for a summary). Those over 70 years of age showed larger longitudinal declines than those between the ages of 30 and 70 for all tests.

Gilbert (1973) reported reliable declines on initial learning and retention subtests but no changes on easy–old material (combined scores from questions on general information, including naming the months) or repetitions (a combined measure including digit span, spatial span, and sentence repetitions) over 35–40 years in 14 individuals. McCarty, Siegler, and Logue (1982) analyzed data from the Duke Longitudinal Study and reported 4- and 9–11-year declines on visual memory and the hard associates subtests of the Wechsler Memory Scale (Wechsler, 1945) but no changes in easy associates or immediate and delayed logical memory (prose recall). A subset of 26 subjects with 16-year longitudinal data showed declines only on the visual memory test.

Why are the group longitudinal studies inconsistent in findings? One possibility is that there may be power problems: Studies with smaller samples have mixed results, whereas the BLSA, with over 200 participants, does not. Another is that in very-long-term studies, only those who have excellent memory abilities continue as participants. A third possibility is that there are differences across tasks in their sensitivity to detect memory change, because tasks also vary in their sensitivity to detect cross-sectional age differences (e.g., Light, 1991).

## Cross-Sectional and Longitudinal Individual Differences in Memory and Aging

In the present research, we examined the question of how individual differences account for memory performance across the adult life span. Two major tasks were used: recall of a list of unrelated concrete nouns and prose recall. These probably represent different memory demands, because it has been shown that cross-sectional and longitudinal differences are likely to occur under unrelated list conditions, whereas differences of equal magnitude with a prose recall task may be less likely (e.g., McCarty et al., 1982). On the other hand, it is possible, as Hultsch et al. (1990) found, that approximately the same individual differences factors account for list and prose recall performance.

We present cross-sectional and 3-year longitudinal data from a study of adults aged 55 to 84. The tasks that subjects completed were included as part of a brief memory battery (requiring about 30 min to complete) within a larger study norming the Schaie-Thurstone Adult Mental Abilities Test (STAMAT; Schaie, 1985), a variant of the Primary Mental Abilities Test (PMA; Thurstone & Thurstone, 1949). The PMA took approximately 2.5 hr to complete. Because the memory assessment was secondary to the purpose of the psychometric measurement of intelligence, the memory tasks were limited to prose and list remembering. Subjects also completed the Memory Functioning Questionnaire (MFQ; Gilewski, Zelinski, & Schaie, 1990), a measure of memory self-appraisal, so that the level of subjective memory ability could be examined as a predictor of individual differences. In our previous work, we had found that once the effects of depression, health, and schooling had been partialed, two of the MFQ factor scores accounted for up to 10% of the variance in memory performance (Zelinski et al., 1990).

The cross-sectional data analyses focused on the prediction of individual differences in memory as a function of age, demographic characteristics, reasoning, spatial ability, vocabulary, and memory self-appraisals. Following previous research, we expected to find that the psychometric measures were consistent and reliable predictors of memory. It was not clear whether age would emerge as a reliable predictor under models in which reasoning and vocabulary were entered early in the regression equation. We did, however, expect that memory self-appraisals would account for small but reliable amounts of variance in memory performance as well.

Longitudinal analyses investigated individual differences in patterns of longitudinal performance after partialing the effects of initial performance, along with initial subject background and abilities. We predicted that correlations between longitudinal performance and initial performance would be high, that the amount of variance accounted for by the individual differences variables would also be reliable, and that the longitudinal individual differences measures would account for residual variance in longitudinal performance.

## Cross-Sectional Analyses

### *Method*

*Subjects.* Complete data were available for 508 subjects aged 55–84 from the initial test and, of these, 227 longitudinal subjects. There were approximately equal numbers of men (250) and women (258) at Time 1 (T1). Time 2 (T2) results are reported after the cross-sectional ones.

Subjects were members of a health maintenance organization based in Long Beach, California. Approximately 3,000 individuals from the membership in an age-stratified sample were contacted by letter to volunteer to participate in a 3.5-hr testing session. Of these, 590 actually participated. Of the subjects older than 55 who participated, 349 indicated their socioeconomic status on a scale ranging from *unskilled* (0) to *professional* (9); the mean level was 5.01 ($SD$ = 2.17). Also, 410 subjects rated their health on a scale ranging from *very poor* (1) to *excellent* (10); the mean rating was 7.38 ($SD$ = 1.85). Because not everyone responded to these questions about health and socioeconomic status, these variables were not analyzed in the regressions. Analyses of data from 99 older subjects from the initial sample have been reported elsewhere (Zelinski et al., 1990, Study 1), but their data are included here as well to increase the reliability of findings.

*Materials and procedure.* The memory tasks were those used in our previous research (Zelinski et al., 1990; Zelinski, Gilewski, & Thompson, 1980). The first task involved study of a typed list of 20 concrete one- or two-syllable English nouns for 3.5 min, which was free recalled immediately afterward (immediate list recall). In the second task, subjects read a 227-word essay about parakeets as house pets while listening to a taped reading of the essay at approximately 155 words per minute. They then wrote an immediate free recall of the essay (prose recall). Subjects were encouraged to recall the text verbatim but also to include anything that they remembered in their own words. Another unrelated task followed, and subjects were asked to free recall the words from the list studied earlier (delayed list recall). The delayed list recall task occurred approximately 20 min after the list was first studied. The final memory task, occurring after delayed list recall, was a recognition test involving selection of the 20 words studied earlier from a list including the 20 words and 20 foils, 10 of which were synonyms of 10 words from the original list and 10 of which were unrelated words (list recognition).

Two-week test–retest reliabilities on the memory tasks were obtained from an independent sample of 72 subjects ranging in age from 18 to 76. Prose recall had a reliability of .70; immediate list recall, .76; delayed recall, .72; and recognition, .74.

The MFQ (Gilewski et al., 1990), a 64-item self-report instrument for assessing individuals' perceptions of their memory abilities on four dimensions—Frequency of Forgetting, Seriousness of Forgetting, Retrospective Functioning, and Mnemonics Usage—was also administered. Approximately half the subjects responded to the MFQ before they completed the memory tasks and half after. There were no differences in scoring patterns as a function of when the MFQ was completed. Although no test–retest reliabilities were computed, there is evidence of the MFQ's reliability (see Gilewski et al., 1990).

The recall–recognition phase of all memory tasks was self-paced. The same procedures and order of testing were used for the 3-year retest.

*STAMAT.* The STAMAT (Schaie, 1985) was adapted for use with older populations, with disposable test booklets in which answers were directly marked, as opposed to standard intelligence tasks in which responses are marked on machine-scorable sheets. Items were presented in large type. Although these adaptations do not eliminate age differences due to perceptual and motor speed deficits, they do reduce them (e.g., Hertzog, 1989).

There are parallel forms of tests for the Spatial Ability and Inductive Reasoning tasks, one form with the original items from the PMA and the other with more familiar items.

The specifics of each STAMAT test are as follows. The Letter Series test is an inductive reasoning task in which the subject indicates what the next letter should be in a series (e.g., abccbadeffe). Scoring is of the number correct, and the highest score is 30.

The Word Series test is a parallel form of the Letter Series task in which the subject reads series of words (e.g., January March May) and indicates the next word in the series. Scoring is identical to that of Letter Series.

The Figure Rotation task is a spatial rotation test in which subjects see a line drawing of an abstract figure and select the items from an array corresponding to rotations of the figure in two-dimensional space. Scoring is of the number correct minus the number incorrectly selected, with 70 the highest score.

The Object Rotation test is a parallel form of the Figure Rotation task that uses line drawings of common objects rather than abstract figures. For example, subjects determine which drawings of a bleach bottle are two-dimensional rotations of a standard. Scoring is as for Figure Rotation.

The alternative forms of the reasoning and spatial tasks were counterbalanced in terms of presentation order.

The Recognition Vocabulary test involves selection of the appropriate synonym of a word from a field of four choices. The number of correct synonyms is the score, and the highest score is 50.

The 2-week test–retest correlations of these tests for 172 subjects ranged from .85 to .89 (Schaie, Willis, Hertzog, & Schulenberg, 1987).

*Data scoring.* Prose recall was scored by parsing the passage into 74 content units (propositions) and relations between content units with Meyer's (1975) system of prose analysis. Recall protocols were compared with the analyzed passage and scored by two scorers for the presence of both content units and relations. For a sample of 10 passages, the interscorer reliability was .92.

The proportion of correctly recalled items was the score for the list recall tests, and an index of the probability of correctly recognized items ($d'$) was used as the score for the list recognition test.

For ease of comparison across factors, the mean ratings for Frequency of Forgetting and Seriousness of Forgetting from the MFQ were calculated by summing scores within factors and dividing them by the number of items in each factor. Only the frequency and seriousness factor scores were included in the present analysis to reduce the number of predictors (see Cohen & Cohen, 1975). Also, we have reported elsewhere that these two factor scores are the best predictors of the four MFQ scores for memory performance (Zelinski et al., 1990).

Schaie et al. (1987) have reported that the Letter Series and Word Series tasks load on a reasoning factor; their correlation in our sample was .85, and so the raw scores for these two tasks were added together for a single reasoning score. Similarly, because the two rotation tasks load on a space factor and their correlation was .78 in the present sample, both raw Figure Rotation and Object Rotation scores were summed.

The original PMA version of the STAMAT Recognition Vocabulary test loads on a verbal abilities factor and a perceptual speed factor, with speed a strong predictor of performance in older adults. Hertzog (1989) has shown that once effects of speed are partialed from PMA vocabulary scores, scores increase with age. This suggests that the STAMAT Recognition Vocabulary subtest, with answers recorded directly therein, is much less affected by perceptual speed than its parent test, and we will consider it to be an approximation to a measure of verbal ability.

The means, standard deviations, and intercorrelations for all variables examined in the study are found in Table 1.

*Data analysis.* Hierarchical regression analyses with each of the memory task scores were computed, and fixed orders of variables were entered under two models. Ordering reflected assumptions about the relative importance of factors predicting memory performance. Retest participation was always entered into the equation first because retestees almost invariably perform at higher levels on the initial test than do those who drop out (e.g., Schaie, 1988).

The next sets of predictors were varied in specific sequences. In

Table 1
*Correlation Matrix, Means, and Standard Deviations for the Cross-Sectional Study at Time 1*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Prose | — | | | | | | | | | | | |
| 2. List recall | .44 | — | | | | | | | | | | |
| 3. *d'* | .32 | .64 | — | | | | | | | | | |
| 4. Retest participation[a] | .16 | .16 | .18 | — | | | | | | | | |
| 5. Age | −.24 | −.37 | −.21 | −.02 | — | | | | | | | |
| 6. Gender (250 men) | .16 | .24 | .23 | .01 | .01 | — | | | | | | |
| 7. Schooling | .21 | .27 | .11 | .06 | −.26 | −.09 | — | | | | | |
| 8. Frequency of forgetting | .14 | .13 | .06 | .02 | −.13 | .00 | .15 | — | | | | |
| 9. Seriousness of forgetting | .01 | −.03 | −.08 | .02 | −.00 | .04 | −.02 | .48 | — | | | |
| 10. Reasoning | .43 | .50 | .29 | .18 | −.43 | .06 | .33 | .08 | −.04 | — | | |
| 11. Space | .27 | .25 | .20 | .12 | −.37 | −.25 | .18 | .11 | −.04 | .49 | — | |
| 12. Vocabulary | .41 | .44 | .28 | .18 | −.22 | .05 | .31 | .20 | .03 | .35 | .35 | — |
| *M* | 0.39 | 0.45 | 1.48 | — | 68.71 | — | 12.70 | 4.76 | 4.87 | 19.86 | 38.21 | 37.82 |
| *SD* | 0.19 | 0.18 | 0.98 | — | 7.76 | — | 2.83 | 0.82 | 1.11 | 11.21 | 21.03 | 10.96 |

*Note.* All values of correlations over .11 are significant at $p < .01$.
[a] Two hundred twenty-seven subjects (106 men) were retested.

Model 1, we next examined the main effect of chronological age. After partialing age, the subject background variables of gender and years of schooling were entered because there is evidence that female subjects recall better than male subjects (e.g., Hultsch et al., 1990), and educational attainment is also related to better memory (e.g., Salthouse, 1991). The next blocks of predictors were reasoning, entered first because it is related to learning and memory (Kyllonen & Christal, 1990), followed by space, as another fluid ability, and then vocabulary score, a crystallized ability, the level of which is related to fluid abilities (e.g., Horn, 1968).

The remaining blocks of predictors were entered in a fixed order. The Frequency of Forgetting and Seriousness of Forgetting scores were entered in one block into the regression next because they have been found to predict memory after partialing health, schooling, and affective status (Zelinski et al., 1990) and are themselves independent of age, health, and schooling (Gilewski et al., 1990).

The final block of variables involved terms testing interactions of age with some of the main effects. There is some evidence (Cooney, Schaie, & Willis, 1988) that age may interact with the effects of retest, whereby older adults, whose reasons for dropout may include chronic health problems or their impending demise, show larger retest differences than younger ones, who are more likely to not retest for social reasons, including having moved away from the area. Thus, we tested the Age × Retest interaction as a factor in memory performance. Following Hultsch et al. (1990), interactions of age with the three STAMAT scores were entered to determine whether memory is related to differential effects of ability as a function of age, with less differentiation in older adults.

The method of computing interaction predictors was that proposed by Aiken and West (1991). The main effect variables of age, retest, reasoning, space, and vocabulary were centered to a mean of zero by subtracting the mean value of each variable from each subject's score. Thus, the distribution of scores, and therefore standard deviations and regression coefficients, remained the same. Interaction terms were computed by multiplying the centered scores so that they were unaffected by scale differences across measures. The interaction block was last under both models because interactions cannot be entered into the regressions until the main effects have been partialed (Cohen & Cohen, 1975).

The order of entry of predictors for Model 2 was as follows: retest participation, reasoning, space, vocabulary, gender, schooling, age, the MFQ block, and the interactions.

Comparisons were made of the reliability of the change in $R^2$ across models to examine the unique variance for each predictor block. The level of significance for all analyses was set at $p < .01$ because of the large number of comparisons.

## Results

Because immediate and delayed recall were highly intercorrelated ($r = .89$), the raw scores for both measures were summed and analyzed as list recall. The last row of Table 2 shows that the regression equations accounted for approximately one fifth to one third of the variance in memory scores, $R^2 = .274$ for prose recall, .392 for list recall, and .203 for recognition. All $F$ values were significant, $Fs(13, 494) = 13.73, 24.48$, and 9.75.

*Memory for prose.* As seen in Table 2, retest status accounted for 2.6% of the variance in prose recall; retestees were more likely to recall a greater amount of information. Predictors reliable for both regression models were gender, reasoning, and vocabulary.

Gender accounted for 2.3% of the variance in prose recall for both models, with women outperforming men.

When reasoning scores were entered early in Model 2, the $R^2$ change was .162; when entered after age, gender, and schooling had been partialed, these scores accounted for 8.5% of the variance in prose recall. Vocabulary accounted for 3.1% of the variance in prose recall in Model 1 and 3.7% in Model 2, where it was entered early.

The predictors reliable only before abilities were entered into the regression (i.e., under Model 1) were those regarded to be important predictors of prose recall: age ($R^2$ change = .057) and years of schooling ($R^2$ change = .026). Obviously, variance in prose recall possibly attributable to these variables was redundant with reasoning and vocabulary.

Space was also reliable only under Model 1 (when it was entered late), as a result of cooperative suppression of space and gender (see Cohen & Cohen, 1975). As seen in Table 1, first-order correlations between space and prose recall and gender

Table 2
*Results of Regression of Subject Background Variables on Memory Test Performance*

| Model/variable | Prose recall | | | List recall | | | Recognition | | |
|---|---|---|---|---|---|---|---|---|---|
| | $F$ change | $R^2$ change | Cumulative $R^2$ | $F$ change | $R^2$ change | Cumulative $R^2$ | $F$ change | $R^2$ change | Cumulative $R^2$ |
| Model 1 | | | | | | | | | |
| Retest participation | 13.07* | .026 | .026 | 22.04* | .041 | .041 | 16.03* | .031 | .031 |
| Age | 30.19* | .057 | .083 | 79.85* | .131 | .172 | 22.92* | .041 | .072 |
| Gender | 12.37* | .023 | .106 | 36.01* | .055 | .227 | 29.52* | .051 | .123 |
| Schooling | 14.49* | .026 | .132 | 24.30* | .036 | .263 | 2.43 | .004 | .127 |
| Reasoning | 52.11* | .085 | .218 | 69.45* | .090 | .353 | 18.15* | .030 | .157 |
| Space | 7.18* | .011 | .229 | 1.41 | .001 | .354 | 8.22* | .014 | .171 |
| Vocabulary | 19.90* | .031 | .260 | 20.33* | .026 | .380 | 8.20* | .014 | .185 |
| Model 2 | | | | | | | | | |
| Retest participation | 13.07* | .026 | .026 | 22.04* | .041 | .041 | 16.03* | .031 | .031 |
| Reasoning | 96.27* | .162 | .188 | 153.43* | .224 | .265 | 40.24* | .071 | .102 |
| Space | 2.94 | .005 | .193 | 0.02 | .000 | .265 | 2.14 | .004 | .106 |
| Vocabulary | 23.09* | .037 | .230 | 22.36* | .031 | .296 | 9.01* | .011 | .121 |
| Gender | 14.63* | .023 | .253 | 34.66* | .045 | .341 | 34.56* | .057 | .178 |
| Schooling | 2.76* | .004 | .257 | 8.84* | .012 | .353 | 0.18 | .000 | .178 |
| Age | 1.81 | .003 | .260 | 21.54* | .027 | .380 | 4.30 | .007 | .185 |
| Memory Functioning Questionnaire factor scores | 0.98 | .003 | .263 | 1.64 | .004 | .384 | 2.82 | .009 | .194 |
| Interactions | 1.87 | .011 | .274 | 1.60 | .008 | .392 | 1.46 | .009 | .203 |
| Final equation | 13.73* | | .274 | 24.48* | | .392 | 9.75* | | .203 |

* $p < .01$.

and prose recall are positive, whereas the correlation between space and gender is negative. The net effect is that space and gender symmetrically suppress variance in each other that is irrelevant to prose recall, enhancing each variable's prediction of recall when the other has been partialed out.

Virtually no variance in prose recall in either model was accounted for by the MFQ factor scores or by the interactions of age with retest participation or STAMAT scores.

*List recall.* The reliable predictors across both regression models were age, gender, schooling, reasoning, and vocabulary. Age accounted for 13.1% of the variance in recall when it was entered early and 2.7% when it was entered late. For gender, the $R^2$ change was .055 for Model 1 and .045 for Model 2; for schooling, the $R^2$ change was .036 for Model 1 and .012 for Model 2.

When reasoning entered early, it accounted for 22.4% of the variance; it accounted for 9.0% of the variance in Model 1. Vocabulary accounted for 3.1% of the variance in Model 2 and 2.6% in Model 1. Space was not a significant predictor under either model.

Of the variables with constant positions across both models, only retest, which accounted for 4.1% of the variance in list recall, was reliable. Neither MFQ factor scores nor the interactions added reliably to the equations.

*Recognition.* As in the other analyses, gender, reasoning, and vocabulary were reliable predictors across both models. When entered in Model 1, the $R^2$ change for age was .041; when entered late, in Model 2, it was .007 and not reliable. The corresponding values for gender were .055 and .057.

For reasoning, the contributions to $R^2$ for recognition were .071 in Model 2 and .030 in Model 1. Vocabulary contributed 1.1% in Model 2 and 1.4% in Model 1. Space was a reliable predictor only in Model 1, showing cooperative suppression with gender.

Of the variables in constant positions, retest participation had a reliable $R^2$ of .031. The MFQ block was marginal, with $F$ significant at $p < .06$ and the interaction block not reliable.

## Discussion

*Overall findings.* It is clear from the results presented here that manipulating the position of selected predictors in hierarchical regression models produces variations in the size of their $R^2$s. This supports the point that when variables are highly intercorrelated, as are age and reasoning, step-down approaches, as used by Hultsch et al. (1990), do not fully address questions about the independent contributions of abilities and age to memory performance because they do not identify the unique contribution of each of the critical predictors.

Our findings do, however, confirm the results of Hultsch et al. (1990) and Cockburn and Smith (1991) in revealing that individual differences in intellectual abilities in adults of different ages mediate memory performance. Fluid abilities, as suggested by Rabbitt (1989), are excellent (and, in our study, the best) predictors of memory performance. We also found that vocabulary and gender were consistent predictors. Age and schooling, in addition, predicted list recall. Thus, there are multiple predictors of memory performance.

In general, none of the variables accounts for especially large amounts of variance in memory, but reasoning, age, and gender account for most of the variance in the tasks studied. Reasoning accounts for more variance in memory than age for the prose recall and recognition measures because it was a reliable predictor under both models, whereas age was not. Reasoning also accounted for more variance in list recall when entered early (22.4% vs. 13.1%) as well as when entered late (9.0% vs. 2.7%). Gender generally accounted for approximately 5% of the variance in all three memory measures, and vocabulary accounted for 1% to 3%, regardless of their positions in the regression models.

The theoretical ramifications of these findings are perhaps less clear than originally assumed. The relatively strong relationship of reasoning with memory scores suggests that the mechanisms of reasoning with respect to memory need to be identified to understand memory performance in aging. On the other hand, the findings for list recall also support a model of mechanisms underlying age differences after the main effects of reasoning, vocabulary, gender, and schooling have been partialed, because age was a reliable predictor under both models. It may seem that age is a rather weak predictor, relative to reasoning, but one must keep in mind the age range of the sample studied here, with the youngest subjects at the age of 55. If we had included the data of young subjects, age would have been a reliable predictor for all memory measures.

Of the predictors entered in positions constant in both models, only retest was reliable. It accounted for about 3% of the variance in performance on the three memory measures. Given that it was entered first into the regressions, it seems that the effects of retest in longitudinal research are not as crucial as may have been assumed. Furthermore, computing the regression equations without including retest resulted in virtually the same amounts of variance accounted for by the other predictors, indicating that the predictors of memory performance are independent of retest status. Thus, those who remain in longitudinal research may perform better on memory measures at T1 than those who do not; however, reasoning, vocabulary, age, gender, and schooling are much more important predictors.

Memory self-appraisals did not predict memory performance in the present analysis, and this is contradictory to our previous research (Zelinski et al., 1990). However, in the present case the predictive power of MFQ scores, which is not substantial to begin with, is probably redundant with age and abilities, which were not partialed in previous research; also, because these scores entered the regressions very late, there was little additional variance in memory scores that they could have accounted for. We tested this conclusion by entering the block of MFQ scores first in the regression analyses; this produced small but reliable $R^2$s of .023, .028, and .018 for prose recall, list recall, and recognition, respectively.

We also did not obtain any reliable interactions of retest participation or reasoning, space, and vocabulary with age. This suggests that specific effects of individual differences do not have differential effects in adults of different ages. Hultsch and his colleagues (Hultsch et al., 1990) also found minimal prediction of age interactions in accounting for variance.

*Memory for prose.* In prose recall, the regression findings support a model of reasoning, vocabulary, and gender on performance, with variance attributable to age redundant with reasoning. These findings contrast with the studies of Hartley (1986) and Cockburn and Smith (1991) and may reflect more stable data, because the sample size was fairly large and the predictor-to-subjects ratio relatively small, as compared with those studies. As did Hultsch et al. (1990), we found that intellectual abilities were reliable predictors; however, Hultsch et al. also reported that age independently predicted variance in prose recall. Methodological differences between our study and theirs, including the age ranges of the populations sampled and materials used to assess text recall, probably account for this discrepancy. Our results suggest that to understand prose recall, mechanisms of change relevant to abilities are the underlying sources of age differences. This conclusion confirms to some degree the suggestion of Rice and Meyer (1986) that memory for prose in older adults is better predicted by vocabulary ability than it is by age (see also Zelinski & Gilewski, 1988). However, we found that although vocabulary was a consistent predictor, the best predictor of prose recall was reasoning. These results also support the findings that gender leads to individual differences in prose recall, with women performing at higher levels than men (Hultsch et al., 1990).

*List recall.* In list recall, findings that age was one of the independent covariates of list memory performance contradict those of Hultsch et al. (1990), who found no reliable contribution of age to the prediction of list recall after partialing intellectual abilities. Age effects in our sample involve observations based on a narrower and older age range, thus reducing the likelihood of finding reliable age-related variance in memory in our study; however, the nature of the lists may be the basis of the discrepancies in findings, because Hultsch et al. used categorized lists and we used an uncategorized one. There is some suggestion that categorized lists, with their inherent structure for use of mnemonics, may reduce the variance due to age differences because many younger and some older adults will use strategies to study them; thus, the intellectual abilities Hultsch et al. tested would better predict performance than age alone under those conditions. On the other hand, recall of uncategorized word lists may, as suggested in the longitudinal literature, be sensitive to aging effects independent of individual differences.

*Recognition.* Previous studies have not evaluated the effects of individual differences on recognition. The consistent predictors for recognition were also reliable for prose recall, which suggests that similar processes and sources of individual differences play a role in the level of $d'$. Use of other materials and tasks for recognition in an individual differences study would confirm the reliability of our findings.

## Longitudinal Analyses

We examined the patterns of T2 performance of the 227 people with complete data on all memory and individual difference measures across the two testings. There were 106 men and 121 women originally aged 55–84, and scores over both testing occasions are presented in Table 3.

The method of data analysis was to examine T2 memory performance on each of the three tasks regressed on initial performance for each task, followed by blocks of variables from the T1 measurement in the same orders for Model 1 and Model 2 in the cross-sectional analyses and by blocks of the T2 scores (reasoning, space, vocabulary, and MFQ). No interactions were tested in these models because of the large number of predictors relative to the sample size. This analysis is essentially a

Table 3
*Correlation Matrix, Means, and Standard Deviations for the Longitudinal Sample*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. T1 prose | — | | | | | | | | | | | | | | | | | | |
| 2. T1 list recall | .35 | — | | | | | | | | | | | | | | | | | |
| 3. T1 *d'* | .28 | .64 | — | | | | | | | | | | | | | | | | |
| 4. T2 prose | .50 | .50 | .40 | — | | | | | | | | | | | | | | | |
| 5. T2 list recall | .39 | .72 | .58 | .54 | — | | | | | | | | | | | | | | |
| 6. T2 *d'* | .24 | .55 | .52 | .40 | .70 | — | | | | | | | | | | | | | |
| 7. Age | −.14 | −.41 | −.22 | −.26 | −.42 | −.29 | — | | | | | | | | | | | | |
| 8. Gender (106 men) | .11 | .23 | .23 | .10 | .22 | .21 | .10 | — | | | | | | | | | | | |
| 9. Schooling | .18 | .26 | .04 | .27 | .28 | .18 | −.32 | −.12 | — | | | | | | | | | | |
| 10. T1 Frequency of forgetting | .15 | .09 | .03 | .14 | .11 | .02 | −.18 | −.01 | .12 | — | | | | | | | | | |
| 11. T1 Seriousness of forgetting | −.02 | −.06 | −.17 | −.01 | −.11 | −.15 | −.09 | −.03 | .04 | .48 | — | | | | | | | | |
| 12. T1 reasoning | .35 | .46 | .29 | .38 | .45 | .30 | −.40 | .01 | .37 | .05 | −.02 | — | | | | | | | |
| 13. T1 space | .21 | .31 | .27 | .32 | .29 | .22 | −.40 | −.30 | .19 | .10 | −.01 | .48 | — | | | | | | |
| 14. T1 vocabulary | .33 | .44 | .26 | .38 | .35 | .24 | −.21 | .03 | .34 | .17 | .10 | .52 | .32 | — | | | | | |
| 15. T2 Frequency of forgetting | .13 | .21 | .02 | .20 | .21 | .13 | −.16 | .04 | .11 | .63 | .28 | .11 | .09 | .19 | — | | | | |
| 16. T2 Seriousness of forgetting | −.02 | −.02 | −.15 | .02 | −.01 | −.08 | .00 | .05 | .01 | .25 | .44 | −.04 | −.06 | .01 | .39 | — | | | |
| 17. T2 reasoning | .35 | .46 | .30 | .35 | .45 | .30 | −.49 | .04 | .36 | .08 | .00 | .87 | .52 | .47 | .14 | .00 | — | | |
| 18. T2 space | .29 | .36 | .30 | .48 | .34 | .24 | −.54 | −.25 | .27 | .17 | .08 | .54 | .84 | .36 | .09 | −.06 | .61 | — | |
| 19. T2 vocabulary | .36 | .39 | .27 | .43 | .35 | .31 | −.27 | .01 | .32 | .09 | .05 | .56 | .35 | .84 | .20 | .04 | .53 | .42 | — |
| *M* | 0.43 | 0.49 | 1.67 | 0.44 | 0.51 | 1.62 | 68.56 | — | 12.92 | 4.77 | 4.89 | 22.22 | 40.54 | 40.07 | 4.59 | 4.66 | 21.82 | 40.54 | 40.20 |
| *SD* | 0.17 | 0.16 | 1.05 | 0.17 | 0.19 | 1.06 | 7.62 | — | 2.78 | 0.77 | 1.11 | 10.10 | 21.80 | 9.98 | 0.80 | 1.14 | 10.42 | 21.80 | 9.82 |

*Note.* All values over .16 are significant at $p < .01$. T1 = Time 1; T2 = Time 2.

partial analysis of covariance whereby the effects of the T1 predictors, including T1 memory scores, have been partialed out before the T2 predictors have entered the equation (Cohen & Cohen, 1975). This method reduces problems in reliabilities of predictors as well as of change scores while examining whether there are consistent patterns of change, because the correlates of T1 levels of performance have been partialed out. Thus, we were able to determine whether changes in STAMAT scores as well as memory self-assessment were predictive of changes in memory scores.

The results seen in Table 4 indicate that the final regression equations accounted for reliable amounts of variance, $F$s(12, 214) = 8.75, 21.64, and 8.64 and $R^2$s = .384, .593, and .369 for prose recall, list recall, and recognition, respectively. As expected, most of the variance in T2 performance (24.6%, 52.2%, and 26.6%, respectively) was accounted for by T1 performance.

Once the T1 score had been partialed, approximately 7% to 13% additional variance in T2 performance was accounted for by individual difference variables, but none was consistently reliable at $p < .01$ across the two models tested.

### *Prose Recall*

Under Model 1, age was entered after T1 score, and it accounted for 3.8% of the variance. The T1 space score reliably accounted for 2.3%, showing cooperative suppression with gender. Under Model 2, T1 reasoning, when entered early, accounted for 5% of the residualized variance in prose recall.

### *List Recall*

In Model 1, age reliably accounted for 1.6% of the variance; in Model 2, T1 reasoning accounted for 1.7%. The T2 reasoning scores were always entered in the same position for both models and reliably accounted for another 1.6% of the variance in list recall.

### *Recognition*

Here the $R^2$ associated with age in Model 1 was .033; for reasoning, in Model 2, the $R^2$ was .024. No other measures had reliable prediction.

### *Discussion*

With T1 memory performance and all other T1 measures having been partialed, the T2 scores accounted for little additional variance. For list recall, the additional 1.6% contribution of T2 reasoning was reliable, however. The positive nature of the relationship between T2 reasoning and list recall suggests that negative residualized list recall scores (i.e., lower than predicted T2 recall) were accompanied by parallel changes in reasoning. Thus, individuals who declined in delayed recall also declined in reasoning, and vice versa.

The findings for prose recall and list recognition indicate that T2 performance, once the effects of T1 predictors are partialed out, is not predicted by longitudinal score on intellectual abilities, contrary to our predictions. The reason may be that the T1 scores of the predictors of memory performance were highly correlated with the T2 scores, as seen in Table 3, so the

Table 4
*Individual Differences in Time 2 Performance*

| Model/variable | Prose recall F change | Prose recall $R^2$ change | Prose recall Cumulative $R^2$ | List recall F change | List recall $R^2$ change | List recall Cumulative $R^2$ | Recognition F change | Recognition $R^2$ change | Recognition Cumulative $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 | | | | | | | | | |
| T1 performance | 68.42* | .246 | .246 | 240.08* | .522 | .522 | 79.65* | .266 | .266 |
| Age | 11.22* | .038 | .284 | 7.77* | .016 | .538 | 10.28* | .033 | .299 |
| Gender | 1.53 | .005 | .289 | 3.94 | .008 | .546 | 4.66 | .015 | .313 |
| Schooling | 5.54 | .018 | .307 | 3.00 | .006 | .552 | 4.21 | .013 | .326 |
| T1 reasoning | 5.34 | .018 | .325 | 3.97 | .008 | .560 | 1.48 | .005 | .331 |
| T1 space | 7.24* | .023 | .348 | .062 | .001 | .561 | 0.45 | .001 | .332 |
| T1 vocabulary | 3.92 | .031 | .361 | .013 | .001 | .562 | 0.06 | .000 | .332 |
| Model 2 | | | | | | | | | |
| T1 performance | 68.42* | .246 | .246 | 240.08* | .522 | .522 | 79.65* | .266 | .266 |
| T1 reasoning | 15.02* | .050 | .296 | 8.03* | .017 | .539 | 7.57* | .024 | .290 |
| T1 space | 5.69 | .019 | .315 | 0.13 | .000 | .539 | 0.08 | .000 | .290 |
| T1 vocabulary | 5.57 | .018 | .333 | 0.12 | .000 | .539 | 0.33 | .001 | .291 |
| Gender | 3.53 | .011 | .344 | 3.71 | .008 | .547 | 4.15 | .013 | .304 |
| Schooling | 3.63 | .011 | .355 | 3.06 | .006 | .553 | 4.26 | .014 | .318 |
| Age | 1.59 | .006 | .361 | 4.28 | .009 | .562 | 4.52 | .014 | .332 |
| T1 MFQ factor scores | 0.28 | .002 | .363 | 2.88 | .012 | .574 | 1.46 | .009 | .341 |
| T2 reasoning | 1.90 | .006 | .369 | 8.26* | .016 | .590 | 0.14 | .000 | .341 |
| T2 space | 0.06 | .000 | .369 | 0.23 | .000 | .590 | 0.38 | .001 | .342 |
| T2 vocabulary | 3.08 | .010 | .379 | 1.73 | .003 | .593 | 4.82 | .015 | .357 |
| T2 MFQ factor scores | 0.82 | .005 | .384 | 0.10 | .000 | .593 | 1.73 | .011 | .368 |
| Final equation | 8.75* | | .384 | 21.64* | | .593 | 8.64* | | .368 |

*Note.* T1 = Time 1; T2 = Time 2; MFQ = Memory Functioning Questionnaire.
* $p < .01$.

likelihood of there being additional variance accounted for by the T2 scores, once the T1 score relationships had been partialed, was not high to begin with.

A second possible explanation is that, given the stability of scores at the group level across testings, as shown in Table 3, individual change may have been unsystematic, leaving little variance that could be predicted. As suggested earlier, it may be very difficult to observe longitudinal change within individuals for psychometric reasons involving reliability within and across occasions (e.g., Salthouse et al., 1986) and because people who experience serious memory declines may drop out of longitudinal research.

To test the question of the reliability of T2 change, we reanalyzed the data of those who showed reliable declines or improvements in memory functioning. Subjects were so classified if their across-measurement difference scores were ±1 standard error of the T1 sample error (see Schaie, 1989, for a description). Although objections have been raised to this approach because it uses between-subjects error as an estimate of within-subjects error (Salthouse, 1991), it does provide an objective criterion of change relative to the initial level of performance. There were 82 individuals showing reliable change on prose recall, 75 on list recall, and 41 on recognition. Of these, the number declining and the number improving were 37 and 45 for prose recall, 28 and 47 for list recall, and 24 and 21 for list recognition. Because so few subjects showed change in recognition scores, there is probably considerable unreliability in the $d'$ scores; thus, we will disregard those findings.

In this set of analyses, the T1 scores were the dependent variables. Because there had been so many predictors in the main longitudinal analyses and our subsamples were small, we included only four predictors: the initial memory score for each measure, the subjects' age at T1, and the reasoning scores from T1 and T2. These variables were chosen because of their relevance to the hypotheses about age and reasoning tested in this article. There were two orders of predictors: (a) the T1 memory score, then age, T1 reasoning, and T2 reasoning for the first model, and (b) the T1 memory score, T1 reasoning, age, and then T2 reasoning for the second model.

We found that T1 scores accounted for reliable variance in prose recall ($R^2 = .55$), with a negative beta indicating regression to the mean (those with high T1 scores declined, whereas those with low T1 scores improved). Once this was partialed, however, age accounted reliably for an additional 5% of the variance in prose recall for Model 1, whereas reasoning accounted for 2%, which was not reliable. For Model 2, the pattern of reliability switched, with the 4% of the variance accounted for by T1 reasoning reliable and the 2% by age not reliable. For list recall, there were no effects of regression to the mean (because immediate and delayed recall scores had been summed) and no reliable prediction of T1 scores. However, age was a significant predictor, accounting for 10% of the variance in Model 1 and 7% in Model 2. The T1 reasoning score was not a reliable predictor of longitudinal performance for either model.

These results suggest that even when regression to the mean exists across testing occasions, additional variance in change is explained by individual difference variables. In prose recall, the pattern seen for the analysis including all subjects was replicated: Depending on position of entry into the regression equation, either age or reasoning was reliable. In list recall, the pattern differed: Only age was reliable as a predictor in both analyses.

The relationship between age and residualized T2 prose and list recall scores was negative, indicating that declines are most likely to occur in the oldest subjects in our sample, whereas improvement is likely in the younger ones. The relationship between reasoning and residualized prose memory scores was positive, indicating that those with lower reasoning scores declined.

For prose recall, reasoning, which was important in predicting initial performance cross sectionally, was a longitudinal predictor. Reasoning and age were reliable only when entered early in the regressions, which suggests that they share variance due to some common underlying factor but do not account for variance uniquely in prose recall.

With respect to list recall, chronological age was a reliable predictor, regardless of position, after T1 performance was partialed out. This reinforces the notion that age remains a significant source of variance in longitudinal performance in list recall, especially given that the age range for the population studied has been considered "old" by many cognitive aging psychologists. Thus, the longitudinal results suggest that an age change model explains list memory change and that mechanisms relevant to aging rather than reasoning be studied in understanding individual patterns of change.

## General Discussion

The results of the present analyses indicate that individual differences factors predict memory performance on standard laboratory tasks. Such individual differences are as important as age, if not more so, in predicting performance. These findings confirm that, although statistically difficult to separate, there is evidence that it is not merely a given ability such as reasoning or deficits in cognition due to aging that is responsible for poorer memory performance in older than younger adults. As Light (1991) has pointed out, no one hypothesis regarding the nature of age-related differences has emerged as an accurate model of mechanisms in memory that change with age. Indeed, the present findings may suggest that the search for universal mechanisms of change may be somewhat misdirected, given that an array of variables here, as well as in other studies, serve as reliable predictors of up to a third of the variance in concurrent memory performance.

No single study to date has definitively established which individual differences factors play the most critical roles in memory in older adults. However, it is clear from our findings and those of others that reliable predictors include constructs related to reasoning or to fluid intelligence, working memory capacity, speed, verbal ability, age, education, and gender (Cockburn & Smith, 1991; Hartley, 1986; Hultsch et al., 1990).

### *Abilities and Memory*

Relatively little has been published on the issue of why reasoning and vocabulary are such good predictors of cross-sectional memory performance (see also Rabbitt, 1989). We can speculate about several possibilities. One is that reasoning represents a higher order *g* factor (Larson & Saccuzzo, 1989), and, as such, high levels of intelligence are due to good memory ability. Being able to hold a number of items from a reasoning task in memory makes it easier to educe the pattern of their relationships. On the other hand, it may be abstract reasoning ability that helps people to identify and use effective strategies to remember. As we have suggested earlier, reasoning and memory probably interact in adults and produce performance levels based on their synergy.

Although not as good as predictor as reasoning, vocabulary (a measure of crystallized ability) or knowledge consistently accounted for up to 3% of the variance in memory scores. As with reasoning, the direction of the relationship with memory in terms of which is the more primitive cognitive function is not really clear. People high in the ability to learn initially the meanings of difficult words and to retrieve those meanings may have an advantage in creating a mental model representing texts and retrieving them or encoding and retrieving word lists. Alternately, those with high verbal skills may be able to hone memory skills by more frequent practice in memory-enhancing activities such as reviewing and discussing what they have read (Zelinski & Gilewski, 1988). Again, the relationship between verbal ability and memory in adults is probably interactive.

It is also quite probable that other sources of the relationship between ability and memory in older adults include what have been termed processing resources, such as speed or working memory capacity (Salthouse, 1991). More extensive analyses of the nature of these resources, which are highly intercorrelated with each other and with ability, might elucidate the sources of ability differences that can explain why older adults remember less than younger ones (see also Hultsch et al., 1990).

### *Age and Memory*

In the cross-sectional analysis, age was a consistent predictor of list memory, although not as strongly as were reasoning and vocabulary. In terms of prediction of T2 performance for those who experienced change by our criteria, however, age predicted residualized list recall, whereas reasoning did not. In longitudinal prose recall, performance was equally predicted by age and reasoning. This suggests that some processes we have not identified in the present study predict change in memory with age over the brief period of 3 years. It is not known whether these changes involve some generic biologically based process, cognitive processing efficiency, or social processes or interact with health or distance from death (see Johannson & Berg, 1989).

Because virtually no other studies have examined longitudinal patterns of individual change in memory, the discrepancies between our longitudinal findings, which emphasize age effects, and the cross-sectional results, which also emphasize reasoning, gender, and vocabulary, await resolution in future research. Although our data represent an initial step in examining the roles of age, reasoning, and other mechanisms of memory and how they may interact to produce the effects observed, in some ways they have raised more questions than we have been able to answer. Cross-sectional and longitudinal list recall follow a model of changes with age once individual differences have been partialed. Cross-sectional prose recall and list recognition and longitudinal prose recall follow a model of ability that explains variance in concurrent performance, but age and reasoning share common variance in explaining change. We did not find sufficiently reliable change in recognition to be confident in reporting findings.

These findings will need confirmation from convergent findings from research that includes multiple longitudinal testings to ensure stability of change, as well as examination of a wide variety of possible individual differences mechanisms of cognitive abilities and different memory processes.

## References


Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Arenberg, D. (1983). Memory and learning do decline late in life. In J. E. Birren, J. M. A. Munnichs, H. Thomae, & M. Marios (Eds.), *Aging: A challenge to science and society, Vol. 3, Behavioural sciences and conclusions.* New York: Oxford University Press.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In P. B. Baltes & J. R. Nesselroade (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). San Diego, CA: Academic Press.

Cockburn, J., & Smith, P. T. (1991). The relative influence of intelligence and age on everyday memory. *Journal of Gerontology: Psychological Sciences, 46,* P31–P36.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cooney, T. M., Schaie, K. W., & Willis, S. L. (1988). The relationship between prior functioning on cognitive and personality variables and subject attrition in psychological research. *Journal of Gerontology: Psychological Sciences, 43,* P12–P17.

Crook, T., Bartus, R. T., Ferris, S. H., Whitehouse, P., Cohen, G. D., & Gershon, S. (1986). Age-associated memory impairment: Proposed diagnostic criteria and measures of clinical change—Report of a National Institute of Mental Work group. *Developmental Neuropsychology, 2,* 261–276.

Dwyan, J., & Jacoby, L. L. (1990). Effects of aging and source monitoring: Differences in susceptibility to false fame. *Psychology and Aging, 5,* 379–387.

Gilbert, J. G. (1973). Thirty-five year follow-up study of intellectual functioning. *Journal of Gerontology, 28,* 68–72.

Gilewski, M. J., Zelinski, E. M., & Schaie, K. W. (1990). The Memory Functioning Questionnaire for assessment of memory complaints in adulthood and old age. *Psychology and Aging, 5,* 482–490.

Hartley, J. T. (1986). Reader and text variables as determinants of discourse memory in adulthood. *Psychology and Aging, 1,* 150–158.

Hertzog, C. (1989). Influences of cognitive slowing on age differences in intelligence. *Developmental Psychology, 25,* 636–651.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75,* 242–259.

Hultsch, D. F., & Dixon, R. A. (1990). Learning and memory in aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 258–274). San Diego, CA: Academic Press.

Hultsch, D. F., Hertzog, C., & Dixon, R. A. (1990). Ability correlates of memory performance in adulthood and aging. *Psychology and Aging, 5,* 356–368.

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review, 85,* 109–130.

Johannson, B., & Berg, S. (1989). The robustness of the terminal decline phenomenon: Longitudinal data from the digit-span memory test. *Journal of Gerontology: Psychological Sciences, 44,* P184–P186.

Kral, V. A. (1962). Senescent forgetfulness: Benign and malignant. *Journal of the Canadian Medical Association, 86,* 257–260.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence, 14,* 389–433.

Larson, G. E., & Saccuzzo, D. P. (1989). Cognitive correlates of general intelligence: Toward a process theory of *g. Intelligence, 13,* 5–31.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42,* 333–376.

McCarty, S. M., Siegler, I. C., & Logue, P. E. (1982). Cross-sectional and longitudinal patterns of memory scores. *Journal of Gerontology, 37,* 169–175.

Meyer, B. J. F. (1975). *The organization of prose and its effect on recall.* Amsterdam: Elsevier/North-Holland.

Rabbitt, P. M. A. (1989). Inner-city decay? Age changes in structure and process in recall of familiar topographical information. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 284–299). Cambridge, England: Cambridge University Press.

Rice, G. E., & Meyer, B. J. F. (1986). Prose recall: Effects of aging, verbal ability, and reading behavior. *Journal of Gerontology, 41,* 469–480.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Erlbaum.

Salthouse, T. A., Kausler, D. H., & Saults, J. S. (1986). Groups versus individuals as the comparison unit in cognitive aging research. *Developmental Neuropsychology, 2,* 363–372.

Schaie, K. W. (1985). *Schaie-Thurstone Adult Mental Abilities Test.* Palo Alto, CA: Consulting Psychologists Press.

Schaie, K. W. (1988). Internal validity threats in studies of adult cognitive development. In M. L. Howe & C. J. Brainerd (Eds.), *Cognitive development in adulthood* (pp. 241–272). New York: Springer-Verlag.

Schaie, K. W. (1989). Perceptual speed in adulthood: Cross-sectional and longitudinal studies. *Psychology and Aging, 4,* 443–453.

Schaie, K. W., Willis, S. L., Hertzog, C., & Schulenberg, J. E. (1987). Effects of cognitive training upon primary mental ability structure. *Psychology and Aging, 2,* 233–242.

Stankov, L. (1988). Aging, attention, and intelligence. *Psychology and Aging, 3,* 59–74.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge, England: Cambridge University Press.

Thurstone, L. L., & Thurstone, T. G. (1949). *Examiner manual for the SRA Primary Mental Abilities Test (Form 10-14).* Chicago: Science Research Associates.

Wechsler, D. A. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19,* 87–95.

Wechsler, D. A. (1958). *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117,* 319–331.

Zelinski, E. M., & Gilewski, M. J. (1988). Memory for prose and aging: A meta-analysis. In M. L. Howe & C. J. Brainerd (Eds.), *Cognitive development in adulthood* (pp. 133–158). New York: Springer-Verlag.

Zelinski, E. M., Gilewski, M. J., & Anthony-Bergstone, C. R. (1990). Memory Functioning Questionnaire: Concurrent validity with memory performance and self-reported memory failures. *Psychology and Aging, 5,* 388–399.

Zelinski, E. M., Gilewski, M. J., & Thompson, L. W. (1980). Do laboratory tests relate to self-assessment of memory ability in the young and old? In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging: Proceedings of the George A. Talland Memorial Conference* (pp. 519–544). Hillsdale, NJ: Erlbaum.


Received March 10, 1992
Revision received October 6, 1992
Accepted December 14, 1992 ■