Memory Aging as a General Phenomenon: Episodic Recall of Older Adults is a Function of Episodic Recall of Young Adults

Paul Verhaeghen and Alfons Marcoen

The focus of this study was on the relationship between young and older adults’ performance on tasks of deliberate recall from episodic memory. A meta-analysis on 91 relevant studies (comprising a total of 154 conditions) was conducted. It was found that 83% of the variance in older adults’ recall probability was accounted for by a quadratic function using young adults’ recall probabilities as predictors. No significant interaction with age of older adults was found. Interaction with task type was, however, significant, resulting in separate functions for separate associate recall. The results point at the importance of the main effect of age in studies on memory aging.

It is a well-established fact that older adults perform less well than do young adults on episodic memory tasks involving deliberate recall. On the average, recall of elderly subjects on typical tasks such as list recall, prose recall, and recall of lists of paired associates is about one standard deviation below that of young adults (Verhaeghen, Marcoen, & Goossens, 1993; standard deviation is standard deviation of the pooled sample). An important issue in current aging research is whether the observed age difference is mainly quantitative or qualitative in nature. On the one hand, if the decrease in episodic memory performance is due to a decrease in a general age-related factor, as is stated in many theories (e.g., Baddeley, 1986; Hasher & Zacks, 1988; Salthouse, 1985, 1988, 1991), one should expect “a systematic relation . . . between the performance levels of young and older adults regardless of the identity or composition of the processing components involved in the task” (Salthouse, 1991, p. 313). If, on the other hand, age differences are much more task specific, for instance, they are due to differences in strategy use (whether caused by deficient knowledge about memory functioning, by disuse, by a production deficiency, by lack of memory self-efficacy, or by monitoring problems; see Light, 1991, for an overview of these explanations), one should expect a less systematic, that is, a more qualitative, relation between the performance levels of both groups.

Unfortunately, very little is known about the relationship between performance levels of different age groups on tasks of deliberate recall from episodic memory. Extensive studies with latency data (Cerella, 1985, 1990; Cerella, Poon, & Williams, 1980; Hale, Myerson, & Wagstaff, 1987; Lima, Hale, & Myerson, 1991; Myerson, Hale, Wagstaff, Poon, & Smith, 1990; and Salthouse, 1991) have demonstrated that latency of the old is indeed a function of latency of the young. This conclusion holds for a wide variety of tasks for which latency data exist, and the proportion of variance accounted for is impressingly large (typically larger than 90%, often exceeding 95%). Linear, quadratic, and power functions, and Myerson et al.’s “information-loss” function all explain the data about equally well. Thus, mean speed of older adults on a wide variety of tasks can be very well predicted simply from knowing the mean speed of young adults on the same tasks; no additional information about the processes involved or the exact nature of the tasks is required. This suggests that speed loss is not task specific but that it is, to a great extent, a general phenomenon.

Very little is known, however, about the relation between young and older adults’ episodic recall performance. In the few studies that have been conducted (Rubin, 1989; Stine & Wingfield, 1988; Verhaeghen & Marcoen, 1993), it was found that the Pearson correlation (after correcting for test unreliability) between the probability that an item is recalled by the young and the probability that the same item is recalled by the elderly of the variance in older adults’ recall probability was accounted for by a quadratic function using young adults’ recall probabilities as predictors. No significant interaction with age of older adults was found. Interaction with task type was, however, significant, resulting in separate functions for separate associate recall. The results point at the importance of the main effect of age in studies on memory aging.

In the present study, we tried to provide a better estimation of
the relationship between young and older adults' episodic recall performance. We conducted a meta-analysis, using conditions within studies as the unit of analysis rather than probability of recall of individual items. All large database on three typical tasks in the episodic memory literature, namely recall of word lists, recall of lists of paired associates, and recall of prose material, was compiled. In this analysis, recall probability of the young, \( p(\text{Rel}Y) \), was related to recall probability of the old, \( p(\text{Rel}O) \), performing identical tasks. Because this database is quite extensive, we trust that the function (if any) relating recall performance of the old to recall performance of the young will give a better estimation of the true relationship between recall performance of the old and recall performance of the young that exists in the population.

Method

The Sample of Studies

For the present analysis, we made use of a database on studies (published from 1975 and after) of adult age differences in episodic (and short-term) memory collected earlier (Verhaeghen, Marcoen, & Goossens, 1993). The database on episodic memory performance consists of 688 studies on list recall, 21 studies on paired-associate recall, and 39 studies on prose recall. These studies were collected through a systematic search of core developmental and gerontological journals (Developmental Psychology, The Gerontologist, International Journal of Aging and Human Development, Journals of Gerontology, and Psychology and Aging). We used references cited in the selected studies and some review papers to identify studies not contained in the original sample. In this database, studies were included that (a) compared a young group (central tendency of age between 16 and 30 years) with a group of older subjects (central tendency of age 60 years or over), both groups being free from cognitive problems caused by organic pathology; and (b) assessed performance on word list recall, prose recall, paired-associate recall (of verbal stimuli), or any combination of these. Selection of studies was concluded in November 1990. From this original data pool, all studies were selected from which probability of recall of word lists, lists of paired associates, and texts could be derived for both age groups. No studies were excluded from the final data pool for reasons other than those stated earlier.

Eventually, we retained 91 studies, contained in 76 articles. These included 53 studies on list recall, 32 on prose recall, and 11 on paired-associate recall. These studies are listed in the Appendix. In total, the memory performance of 3,300 elderly subjects was compared with that of 3,177 young subjects. Mean age (estimated from central tendency measures) was 22.1 years (ranging from 16.9 to 29.6) for the young subgroup and 79.8 years (ranging from 61.4 to 79.8) for the older subgroup.

Statistical Analyses

Statistical analyses were done in two steps. First, data points in the \( p(\text{Rel}Y) \times p(\text{Rel}O) \) space were identified or calculated from available data, for as many between-subject conditions as could be detected in the original articles. To limit the degree of dependence among data points, data from within-subject comparisons within task type (i.e., list recall, prose recall, and paired-associate recall) were averaged. This yielded a fairly extensive database of 81 data points for list recall, 54 data points for prose recall, and 19 data points for paired-associate recall, summing to a total of 154 data points for episodic memory recall. Second, \( p(\text{Rel}O) \) was predicted from \( p(\text{Rel}Y) \), using weighted least-squares analysis, weighing for sample size.

Results and Discussion

A plot of the data points in the \( p(\text{Rel}Y) \times p(\text{Rel}O) \) space for the three types of task can be found in Figures 1, 2, and 3. As can be seen in the figures, the young adults generally outperformed the older adults. The average performance of the elderly was significantly lower than that of the young. Probability of recall for list material was .57 for young adults and .43 for older adults, \( t(80) = 12.67, p < .001 \); for prose recall, probabilities were .41 and .31, respectively, \( t(53) = 10.47, p < .001 \); for paired-associate recall, probabilities were .61 and .41, respectively, \( t(18) = 7.10, p < .001 \). Overall, the young recalled 52% of items in the episodic memory tasks reported here: the elderly recalled 39%, \( t(153) = 16.87, p < .001 \). From the figures, it can also be gathered that recall probability of the elderly only rarely equaled or exceeded that of the young: that is, only 7 data points (9%) for list recall, 3 data points (6%) for prose recall, and 2 data points (11%) for paired-associate recall were situated on or above the diagonal in the \( p(\text{Rel}Y) \times p(\text{Rel}O) \) space. Thus, in the large majority of the conditions in the present sample (92%), the young outperformed the old on episodic memory recall.

Predicting Older Adults' Recall From Episodic Memory

First, a linear function between \( p(\text{Rel}Y) \) and \( p(\text{Rel}O) \) was estimated. This function fitted the data well. \( R^2 = .83, p(\text{Rel}O) = .0778p(\text{Rel}Y) - .0019 \). However, as can be seen in Figure 1, variance of the observations in the data set increased with \( p(\text{Rel}O) \). To test for heteroscedasticity, we applied the method by Goldfield and Quant (1965). In this method, the data are reordered according to the values of the independent variable, and the middle 25% of the observations are deleted (in the present sample, the 39 middle observations were deleted). Two separate regression analyses (i.e., in the present case, least squares analysis, weighted for sample size) are then carried out on the remaining observations, one on the 37.5% observations ranked lowest and one on the 37.5% observations ranked highest. The ratio of the error sum of squares of these regression analyses is \( F \) distributed with the degrees of freedom for both nominator and denominator equal to the degrees of freedom for the error sum of squares in the separate analyses. This statistic should be equal to unity if the error variance is equal for both groups of observations. Another test statistic for heteroscedasticity appropriate when variance is proportional to the values of the independent variable is the rank correlation between the absolute values of the residuals and the values of the independent variable (Kleinbaum, Kupper, & Muller, 1988; for the sake of brevity, this statistic is referred to as rank \( r \) in the remainder of this text). Both of these statistics proved significant. \( F_{\text{high-low}}(57, 57) = 2.01, p < .01; \text{rank } r = .39, p < .001 \), indicating that the variance increased reliably with increases in values of the independent variable. There appears to be no problem with regard to normality of the data (Kolmogorov-Smirnov test on residuals has \( Z = 0.72, n = 97 \)). The heteroscedasticity problem was not resolved after estimation of a quadratic function. This function fitted the data reliably better than did the linear function, \( R^2 = .825, \text{incremental } F(1, 151) = 7.47, p < .01; p(\text{Rel}O) = .0465p(\text{Rel}Y) + .339p(\text{Rel}Y)^2 + .061 \). Residuals were normally distributed (Kolmogorov-Smirnov \( Z = 0.88, n = 97 \)). Hetero-
There were no problems with nonnormality of residuals, but the regression equations, one (but only one) of the two heteroscedasticity statistics was significant. So, probably, interactions with other variables within task type might prove significant.

To solve the heteroscedasticity problem, we included interaction terms in the regression analysis, respectively for the interaction with age of older adults, split at median (69.5 years; note that in three studies, the age of the older subsample was not reported; these studies were omitted from the analysis with an interaction term for age) and with task type (list recall, prose recall, and paired-associate recall). The inclusion of an interaction term for age did not lead to a reliable increase in model fit: linear function with interaction term for age between 3% and 1% of the variance was now significant. So, probably, interactions with other variables within task type might prove significant.

Note, however, that for text recall, heteroscedasticity as indicated by the Goldfield and Quandt (1965) method was now significant in the opposite direction: Variability was larger at the lower end of the $p(Re|Y)$ scale.

An interesting feature of the three quadratic equations presented in the previous paragraph was that the intercept terms in these models proved nonsignificant. This makes sense, because it is reasonable to assume that the regression lines run through the origin, implying that when recall performance of the young equals zero, the elderly will recall nothing as well, whereas the models presented earlier all imply that when recall performance of the young is zero, the elderly will still recall on average between 3% and 21% of the words. Forcing regression lines through the origin resulted in the following equations: (a) for list recall, $p(Re|O) = 0.729p(Re|Y) + 0.008p(Re|Y)^2; R^2 = .790$. $F_{low-high}(28, 29) = 1.13. ns; rank r = .40, p < .001; Kolmogorov-Smirnov Z = 0.65, ns; (b) for prose recall, $p(Re|O) = 0.570p(Re|Y) + 0.366p(Re|Y)^2; R^2 = .907$. $F_{low-high}(19, 17) = 3.31, p < .01; rank r = .12, ns; Kolmogorov-Smirnov Z = 0.86, ns; and (c) for paired-associate recall, $p(Re|O) = 0.154p(Re|Y) + 0.790p(Re|Y)^2; R^2 = .830. F_{high-low}(6.5) = 11.71, p < .05; rank r = -.05, ns; Kolmogorov-Smirnov Z = 0.77. ns. These regression lines are drawn in Figures 1, 2, and 3.
corrected for the mean for the calculation of \( R^2 \) in regression analysis without an intercept term. This results in inflated \( R^2 \) s.

To make the present analysis comparable to the analyses where an intercept term was estimated, we used the total sum of squares corrected for the mean for the calculation of \( R^2 \) values.

Again, there were possibly well-fitted function without heteroscedasticity or transformations overcorrected the problem: reciprocal transformation, fitted the data better (for transformed data: \( R^2 = 0.858 \) and \( 0.789 \), respectively). On the average, quadratic functions with an intercept term (on raw data) explained an extra 2.8% of the variance, and quadratic functions through the origin (on raw data) explained an extra 2.4% of the variance when compared with the transformed data. Moreover, some of the heteroscedasticity statistics signaled possible problems after square root transformation when functions were estimated per task type: for prose recall, rank \( r = -0.31, p < .001 \); for paired-associate recall, \( F_{\text{low-high}}(6, 5) = 8.75, p < .05 \). Thus, data transformation did not result in increased model fit, it did not result in nonsignificance of the interaction with task type, nor did it take away the possible heteroscedasticity problems when regressing within task type.

As a conclusion, it can be stated that mean recall probability of older adults, within task type, can be reasonably well predicted from mean recall probability of younger adults. A quadratic function predicted 83% of the adult age differences in various types of episodic memory functioning. When fitting functions per task type, quadratic functions through the origin explained between 79% and 91% of the variance. However, because of the possible remaining heteroscedasticity, it is likely that there exist further interactions between adult age and characteristics of task and persons (other than age within older subsample) within each task type. In the present study, we have refrained from further investigating possible interactions, first because this was not the focus of the present research and second because in the present state of gerontological science we
have no strong theory that would predict which further interactions might be interesting to investigate.

The fact that the proportion of the variance accounted for is larger in the present analysis as compared with the results from the studies by Rubin (1989), Stine and Wingfield (1988), and Verhaeghen and Marcoen (1993) indicates that some of the scatter in the earlier studies is indeed caused by error components associated with individual items. The proportion of the variance accounted for is, however, smaller than what is typically found in studies on latency differences. At least three reasons might be advanced for this. First, reliability of recall measures may well be smaller than reliability of latency measures. Recall is always measured in a binary way, whereas latency is measured in a continuous way; the binary form of scoring may induce additional scatter. This effect will, however, be rather small with the units of measurement (conditions within studies) used here, compared with the larger unreliability when considering items as units of measurement. Second and more important, recall is an off-line measure; it presupposes encoding, storage, and retrieval of the item. A problem in any of these memory stages may cause a recall failure, and this can, once again, induce additional scatter in the data as compared with on-line measures. There is some evidence that both encoding and retrieval are hampered in old age (e.g., Bayen & Erdfelder, 1992; Howe, 1988). A third reason, of course, is that there may indeed be some qualitative difference; that is, the true relationship may be somewhat weaker than for latency data. A major reason for this may be that there are indeed some differences in strategy use on episodic memory tasks between Young and older adults that influence performance.

Even though the relation between young and older adults’ performance on episodic memory recall may be somewhat weaker than the relationship for latency data, it is still surprisingly strong, considering the diversity in tasks and procedures adopted in the primary studies. For list recall, for instance, some lists involved abstract and others involved concrete nouns; there were conditions of intentional and conditions of incidental learning; there were studies in which recall was tapped immediately or after some delay; there were learning conditions involving mere instruction, or semantic or nonsemantic orienting tasks; recall could be cued or free; and so on. Despite this wide variety of manipulations, presumably involving many different kinds of cognitive processes, 79% of the mean recall performance of the old could be reliably predicted from mean recall performance of the young. The result for mean prose recall performance is even more impressive: 91% of the adult age difference in prose recall performance could be predicted, without any reference to specific characteristics of task or subjects. So, these data offer a very strong case for the hypothesis of a quantitative effect of aging on episodic memory performance. The relationship between recall probability in the young and recall probability in the old appears to be quadratic (somewhat boomerang shaped) rather than linear (note that for list recall, the deviation from linearity is unnoticeable in the figure and in fact not significant). It might be possible that measurement limits (floor and ceiling effects) distort the...
true nature of the relationship, which might then be linear. This remains difficult to test, however (omitting data points below and above specified ranges would also imply that we artificially increase the scatter in the remaining data points, leading to less stable estimates of the function and less statistical power for detecting deviations from linearity).

The Age Main Effect and Consequences for Interactions

When interpreting the data presented here, one mistake should not be made. Strong predictability (between 79% and 91% of the variance accounted for) does not imply that there is no variability or scatter in the data points. There is, quite literally, room for deviations from the regression line. For instance, interactions between adult age and some other variable might still prove significant in primary analyses or meta-analyses. In fact, Fisk, Fisher, and Rogers (1992) recently demonstrated that such interaction effects need not account for large proportions of the variance to become significant. We found an interaction with task type and indications (in our heteroscedasticity analyses) that interactions with other variables might be needed to explain the young-old relationship completely. This is important to note, because much of the research going on in cognitive gerontology is precisely aimed at identifying interactions. Researchers manipulate some variable and look whether age differences are larger in one condition than in another. This research tradition is very valuable and has proven very fruitful. Yet, we think it is important to make two points with regard to this quest for interactions.

First, just as main effects are to be qualified when interactions are present, interaction effects (pointing at specific task-related influences on aging) need to be interpreted in the light of the main effect of age (pointing at a general age-related factor). There are large differences in episodic memory performance between young and older adults; in 92% of the conditions included in the present research, there were age differences favoring the young, and the recall performance from the old can be predicted quite accurately merely from knowing the recall performance of the young on the same task, without making any reference to the type of task or the kind of processes involved. Similar findings for data on young and older adults' performance in latency data have been taken as an indication of a general slowing factor (e.g., Cerella, 1990, 1991; Salthouse, 1991). Likewise, it can be suspected that in episodic recall data a general episodic memory aging factor is at work. This general factor absorbed 83% of the variance, indicating that it is a major determinant of performance. Allowing specific factors for each task type (list recall, prose recall, and Paired-associate recall) through the inclusion of interaction terms resulted in a significant increase of 2.2% of the variance accounted for. So, evidently, α is not the whole picture, but it is an extremely important part of it. It is also crucial to note that this factor is not an explanation of age differences but merely a "descriptive generalization" (Salthouse, 1991, p. 315). It denotes a regularity in a large pool of data, which still needs to be explained by an underlying theory. For instance, for episodic memory recall, explanations relating the decline in older adults to slowing down of cognitive processes have become fashionable, with Salthouse (1985, 1988, 1991; see Light, 1991, for a critique) as the main advocate of this theory.

Of course, focusing on the age main effect, as we do here, does not imply that the quest for specific factors through investigation of interactions is not interesting, worthwhile, or necessary for a better understanding of memory aging. We strongly feel that in theories of memory aging both concepts need to be incorporated, but that perhaps the main effect needs to be explained first, before we can fully understand the nature of interaction effects. Also, precise estimates of the extent of the influence of specific factors can only be made when the general age-related influence has been removed first, to make sure that we are not simply rediscovering a different manifestation of the global factor. As stated most ardently by Salthouse (1992), this "general-plus-specific approach (p. 340) is necessary to achieve a real cumulation of knowledge about causes of adult age differences in cognition.

A second note with regard to interaction effects is a call for caution. The finding that age effects are altered after some experimental manipulation does not always necessarily imply that a local factor is responsible. The shape of the curves computed here implies that age differences will vary with varying task difficulty, regardless of the precise nature of the task involved. Clearly, the relationship between young and older adults' mean recall performance is not additive. So traditional analysis of variance techniques must not be applied to such data (for a related argument concerning latency data, see Nebes & Brady, 1992). For instance, when task difficulty for list recall of young adults is 35% in one experimental condition and 65% in another condition, older adults' recall performance can be expected to be 26% and 48%, respectively. In a classical plot relating condition to recall performance with separate curves for each age group, a divergent interaction along with a main effect for age and manipulation, can be noted. Conversely, when some manipulation increases recall probability from 65% to 75% in young adults for a paired-associate task, the average recall of the old will increase from 43% to 56%, which would amount to a convergent interaction. These observed interactions are not to be explained by a true interaction between age and the nature of the manipulation involved, but merely by the consequences of change in task difficulty. One possible way to counter this problem might be to adopt the method advocated by Cerella (1991), which is to include many conditions, and then estimate both a regression model with main effects only and a regression model with interaction terms, testing for the difference between these two models, or the Fisk et al. (1992) method.

As a general conclusion, it can be stated that probabilities of deliberate recall of verbal material from episodic memory in the old can be predicted to a large extent from recall probabilities of the young. The data suggest that a general aging factor α is at work. This factor seems to be a very powerful contributor to the decline of episodic memory functioning in old age, because it accounts for 83% of the variance in data on older adults' recall. Allowing for specific factors for task types within episodic memory recall increases the proportion of variance accounted for slightly but significantly.

References


Appendix

Studies Included in the Meta-Analysis

(Inc1usion in specific subsamples is indicated by the information between brackets after each reference. LR = list recall; PA = paired-associate recall; PR = prose recall.)


(Appendix continues on next page)


