Age Differences in Mental Multiplication: Evidence for Peripheral But Not Central Decrements

Philip A. Allen, Albert F. Smith, Kathy A. Jerge, and Heli Vires-Collins

Department of Psychology, Cleveland State University.

Mental arithmetic performance is of increasing interest to researchers on aging, in part because there is increasing evidence that older adults retrieve arithmetic information from long-term memory as efficiently as young adults, although older adults appear to take longer to complete peripheral processing (Allen, Ashcraft, & Weber, 1992; Geary, Frensch, & Wiley, 1993; Geary & Wiley, 1991). In fact, Geary et al. (1993) found that older adults borrowed faster in mental subtraction than did young adults. Given that older adults tend to be slower than young adults at almost all information processing tasks (e.g., Cerella, 1985), the finding of an age-related sparing for retrieval time in mental arithmetic deserves careful scrutiny because it suggests that age differences in mental arithmetic performance are process-specific rather than generalized. Advocates of general slowing hypothesize that age differences should be consistent across task components (Cerella, 1985, 1991; Cerella & Hale, 1994; Myerson, Hale, Wragstaff, Poon, & Smith, 1990). However, several studies that have examined age differences in mental arithmetic have concluded that there are age differences in encoding but not in fact retrieval (e.g., Allen et al., 1992; Geary & Wiley, 1991; Geary et al., 1993).

In the mental multiplication tasks of the present study, we examine whether manipulations of the difficulty of central processes (Experiment 1) and of peripheral processes (Experiment 2) differentially influence performance by young and older adults. In a mental arithmetic production task, an individual is asked to answer a presented problem (e.g., 1 × 1 = ?) (see Geary & Wiley, 1991). In a verification task an individual is asked to judge whether an arithmetic statement is "true" or "false" (e.g., 1 × 1 = 1 is true; 2 × 3 = 5 is false) (see Allen et al., 1992). The verification and production tasks may involve different types of retrieval. Namely, the verification task may involve a familiarity judgment rather than the actual fact retrieval process that is presumably required by a production task (Campbell & Tarling, 1996). Experiments employing one or the other of these tasks have failed to find a dependence of retrieval efficiency on age (Geary et al., 1993; Geary & Wiley, 1991; Allen et al., 1992); asking the same individuals to carry out both tasks provides a more sensitive assessment of age differences in fact retrieval. Thus, Experiment 1 included both verification and production tasks.

Experiment 2 examined whether age differences at the encoding stage are present for a mental arithmetic task. Encoding difficulty was manipulated by varying exposure duration in a multiplication production task. The assumption was that decreasing exposure duration would degrade stimulus quality and thereby make encoding more difficult (e.g., Allen, Madden, Weber, & Groth, 1993; Allen, Wallace, & Weber, 1993b). The hypothesis that age-related slowing is specific to encoding would be supported by finding that manipulating stimulus quality affects age differences in encoding performance but has no appreciable effect on age differences in measures of central processes. In contrast, the hypothesis of general slowing would be supported by finding that increasing encoding difficulty increases age differences in measures of central processes (e.g., the Information Loss Model of Myerson et al., 1990).

Measuring Central and Peripheral Processes

Critical to determining whether age differences are generalized (e.g., Cerella, 1985; Myerson et al., 1990) or process specific (Allen et al., 1992; Geary et al., 1993; Geary & Wiley, 1991) is a method of distinguishing among different aspects of processing. In cognitive aging research, perhaps the most fundamental distinction has been between peripheral processes (e.g., encoding and response execution) and central processes (e.g., retrieval and decision processes) (Cerella, 1985, 1991).

The experiments reported in this article use mental arithmetic tasks as a tool to evaluate whether the relationship between age and performance is generalized across processing stage or is process-specific. A pervasive result of studies of arithmetic problem solving is that response times increase...
as a function of the size of an answer; this is known as the problem size effect (Ashcraft & Bataglia, 1978; Groen & Parkman, 1972; Stayzak, Ashcraft, & Hamann, 1982). Regression of response time on problem size yields estimates of the slope—the additional time required per unit increase in answer—and intercept, the response time when the problem size is zero (Allen et al., 1992; Ashcraft, 1992; Geary et al., 1993; Geary & Wiley, 1991). We assume that, in such an analysis, the slope is a measure of central processing efficiency and that the intercept is a measure of peripheral processing speed (see Cerella, 1985, 1991; Geary & Wiley, 1991; also refer to Sternberg, 1967). The analytic strategy pursued in this article, then, is to evaluate the sensitivity of slopes and intercepts of such regression functions to various experimental manipulations and, especially, to the interactions of those manipulations with age.

Fact Retrieval in Mental Arithmetic

The problem size effect is consistent with a network model of associations between problem operands and potential answers stored in long-term memory (e.g., Ashcraft, 1992, 1995; Campbell, 1987; Gallistel & Gelman, 1992; Siegler & Jenkins, 1989). The strength of the connections among operands and answers is assumed to represent the degree of learning (Ashcraft, 1995). According to this network model, speed of fact retrieval is related to the frequency of exposure, the number of competing associations between problems and potential answers, and the absolute value of the problem operands (Ashcraft, 1995; Gallistel & Gelman, 1992). Thus, the model predicts, consistent with the empirical observations, that as problem size increases, longer retrieval time should result.

Another basic empirical observation in mental arithmetic verification research is termed the "split" effect (e.g., Allen et al., 1992; Ashcraft, 1995; Banks, Fuji, & Kayra-Stuart, 1976). Verification problems that are incorrect by a small amount (e.g., 2 x 2 = 5) are more difficult to reject than problems that are incorrect by a large amount (e.g., 2 x 2 = 25). The network model accounts for this finding by supposing that small-split answers are more likely to be associated with the operands (e.g., "1" with "2 x 2") because 2 x 2 = 4) than are large-split answers (e.g., "25" with "2 x 2").

Age Differences in Mental Arithmetic

Charness and Campbell (1988) reported a mental arithmetic skill acquisition study that examined age differences in multiplication and squaring. They concluded that older adults learn and forget how to square numbers within a problem at approximately the same rate as young adults. However, they reported that older adults' calculation speed was about half that of young adults. This study did not separate peripheral (e.g., encoding and response execution) and central (e.g., retrieval) processes. Thus, it was not possible to determine in this study whether encoding and retrieval differed across age groups.

Geary and Wiley (1991) compared the performances of young and older adults on an addition-production task and carried out regression analyses of response times on problem size. They found that there were no age differences in fact retrieval as indexed by the slopes of these regression equations, but they did find that older adults had larger intercepts (a measure of encoding and response production speed; Cerella, 1985, 1991). Geary et al. (1993) replicated these findings using a subtraction-production task. Again, older adults were slower at encoding problems than were young adults, but were as fast as young adults at retrieval. Their study inferred that older adults were actually faster than young adults at the borrowing process of subtraction based on an analysis that estimated the speed of executing that procedure.

Allen et al. (1992) reported data from two multiplication verification experiments. In neither of these experiments were there age differences in problem size or split effects. Regression analyses using problem difficulty as the predictor variable and verification RT as the criterion variable revealed no significant effect of age on slopes. These results led Allen and colleagues to conclude that, even though older adults took significantly longer to respond to problems than did young adults, there were no age differences in fact retrieval for a multiplication verification task.

Salthouse and Coon (1994), however, reported experiments in which the effect size of a problem manipulation depended on age. Salthouse and Coon did not evaluate the relationship of response time and problem size. They used a subtraction verification task in which problems varied according to whether borrowing was necessary, and found that older adults showed a larger borrowing effect than did young adults, a result exactly opposite of that reported by Geary et al. (1993).

Salthouse and Coon (1994) argued that their results differed from those of Geary et al. (1993) because the studies differed in the comparability of the young- and older-adult samples. Specifically, Salthouse and Coon claimed that in their own study, the young and older samples were roughly equal in education, whereas in the Geary et al. study, the sample of older adults was better educated than the sample of young adults. However, this argument seems to be problematic for at least two reasons. First, for many other information processing tasks (e.g., visual search, Allen, Weber, & Madden, 1994), samples of seemingly better-educated older adults have exhibited performance decrements relative to seemingly less-educated young adults. Second, recent cross-cultural studies of mental arithmetic and aging have suggested that older American adults and older Chinese adults do not differ in performance (although older Americans tend toward better performance; Geary, Salthouse, Chen, & Fan, 1996). However, the performance of younger adult Americans was significantly poorer than that of younger Chinese adults. These data suggest that younger American adults, in general, have basic arithmetic skills equal to those of older American adults, but have lagged behind their Chinese age-related counterparts. Of course, the Geary et al. (1996) data could also be interpreted as suggesting that older adults do show a decrement in mental arithmetic (based on the Chinese data), but that the American younger adults are so poorly educated in basic arithmetic that, for Americans, the age deficit is offset by a secular trend in education. The Geary et
ences were in this study. The sampling problem is considera-
bly more complicated than one might surmise from Salthouse and Coon (1994). Consequently, rather than treat-
ing the mental arithmetic phenomenon as a sampling abnor-
mality, we suggest that these data provide evidence that
aging does not necessarily lead to cognitive decline across all
processing stages for certain tasks.

The Present Study

The purpose of the present experiments was to determine
whether age differences in peripheral and central processes
for mental multiplication — to the extent that such differ-
ences exist — are consistent across processing stages (i.e.,
generalized, Cerella, 1985; Myerson et al., 1990) or are
process specific (Allen et al., 1992; Geary et al., 1993;
Geary & Wiley, 1991). The goal of Experiment 1 was to
determine whether age differences across processing stages
varied when retrieval difficulty was varied by manipulating
retrieval task (verification vs production). The goal of Ex-
periment 2 was to test the hypothesis that age differences in
mental arithmetic are due to age differences in encoding, but
not in central processes (Allen et al., 1992; Geary et al.,

Because the questions addressed by these experiments
have been analyzed in a variety of ways in previous relevant
studies, we report several analyses of the data from both of
our experiments even though these are sometimes redundant.
For each study, analysis of variance (ANOVA) is used to
evaluate the dependence of response times and error rates
on the manipulated variables. To obtain indices of central
and peripheral processing performance, the response times
of each participant were regressed on problem size: The
estimated slopes were used as an index of central processing
retrieval task (verification vs production). The goal of Ex-
periment 2 was to test the hypothesis that age differences in
mental arithmetic are due to age differences in encoding, but
not in central processes (Allen et al., 1992; Geary et al.,

The last type of analysis used in this article is Brinley
(plot analysis. In a Brinley analysis, a slowing function
should capture adequately the relationship between mean response times of older and young
adults scored significantly higher (64.6) than older adults
(44.8) on the WAIS-R Digit Symbol Substitution Task
(F[1,38] = 35.91, p < .001), whereas older adults scored
higher (51.9) than young adults (43.0) on the WAIS-R
Vocabulary subscale [F(1,38) = 21.42, p < .001].

Apparatus. — Participants were tested individually using a
Laser 486 (DX2, 50 mhz) computer with an SVGA
monitor and graphics card. Participants’ verification “true”
and “false” responses were collected using the left and right
Materials. — We used the 100 possible multiplication problems with operands between 0 and 9 (i.e., for $a \times b = c$, $a$ and $b$ are the operands), although we report analyses for only those problems with operands between 2 and 9 (in an attempt to ensure that participants were actually retrieving basic arithmetic facts from long-term memory; previous research indicates that problems with operands of 0 and 1 are answered using rules; Ashcraft, 1992). Each of the 100 multiplication problems (64 of which were analyzed) was presented once for the verification task (32 true and 32 false trials were actually analyzed) and once for the production task (i.e., $a \times b = ?$).

For the ANOVA analyses, the problems were assigned to two size categories: small and large. Problems with answers from 4 to 20 were called small; problems with answers from 21 to 81 were called large. Of the 64 problems actually analyzed, 25 problems fit the small problem size criteria and 39 problems fit the large problem size criteria. Previous reports used this definition of problem size (e.g., Allen et al., 1992), so, for continuity, we adhered to the same criterion.

On half of the verification trials, a problem was presented with an incorrect answer. Half of these were small-split trials, on which the presented answer differed from the correct value by either 1 or 2 (e.g., $1 \times 1 = 2$); the other half were large-split trials, on which presented answers differed from the correct value by 3, 4, or 5 (e.g., $1 \times 1 = 6$). For verification problems with a zero operand (e.g., $1 \times 0$), we formed incorrect answers for the small-split condition by using the nonzero operand (e.g., $1 \times 0 = 1$). For verification problems with a zero operand and a large split size, we doubled the nonzero operand (e.g., $1 \times 0 = 2$).

Procedure. — For both the verification and the production tasks, problems were presented horizontally in the center of the computer monitor. Each number subtended approximately 0.28° of horizontal visual angle and 0.56° of vertical visual angle. The average problem subtended approximately 3.30° of visual angle. The stimulus remained on the screen until the participant responded. For the verification task, participants were instructed to decide if the presented problem (e.g., $1 \times 1 = 1$) was true or false; participants responded manually to verification trials by pressing either the left or right arrow key with their right index or middle finger, respectively. For the production task, participants were instructed to answer the presented problem orally (e.g., $1 \times 1 = ?$). Response latencies were computed from the onset of the stimulus to when the participant responded. Task type was blocked, and the order of block presentation was counterbalanced across participants. Instructions gave equal emphasis to speed and accuracy.

RESULTS

Latency data. — Correct response latencies between 100 ms and 3 sec were used in the RT analyses (correct trials outside of this interval were treated as guessing errors; these outliers constituted only 1.96% of the trials). Table 1 presents mean RTs as a function of age, task type, and problem size for problems with operands from 2 to 9.

The 2 (age) x 2 (task type: verification vs production) x 2 (problem size: small vs large) ANOVA revealed that older adults took longer to respond (1237 ms) than young adults

<table>
<thead>
<tr>
<th>Task Type</th>
<th>Verification</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Size</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Small</td>
<td>Young RT</td>
<td>1087</td>
</tr>
<tr>
<td>Older RT</td>
<td>1291</td>
<td>189</td>
</tr>
<tr>
<td>Older % error</td>
<td>9.6</td>
<td>10.8</td>
</tr>
<tr>
<td>Large</td>
<td>Young RT</td>
<td>1143</td>
</tr>
<tr>
<td>Older RT</td>
<td>1393</td>
<td>240</td>
</tr>
<tr>
<td>Older % error</td>
<td>6.9</td>
<td>15.4</td>
</tr>
</tbody>
</table>

Table 1. Reaction Time (RT, in Milliseconds) and Mean Percent Error Data for Problems From Experiment 1 for the Task Type, Verification True-False, and Verification Split Effects Analyses

<table>
<thead>
<tr>
<th>Task Type</th>
<th>Verification: Split Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Size</td>
<td>Small Split</td>
</tr>
<tr>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Small</td>
<td>Young RT</td>
</tr>
<tr>
<td>Older RT</td>
<td>1403</td>
</tr>
<tr>
<td>Older % error</td>
<td>8.8</td>
</tr>
<tr>
<td>Large</td>
<td>Young RT</td>
</tr>
<tr>
<td>Older RT</td>
<td>1475</td>
</tr>
<tr>
<td>Older % error</td>
<td>8.8</td>
</tr>
</tbody>
</table>

For both verification and the production keyboard. Production response times were collected using the Micro Experimental Laboratory (MEL; Schneider, 1988) voice-activated response key connected to a Radio Shack microphone; numerical answers were keyed in by the experimenter. Stimulus presentation and data collection were accomplished using the MEL software; timing accuracy with this software is reported to be accurate within a millisecond (Schneider, 1988).
participants, in general, took longer to respond to verification problems (1229 ms) than to production problems (1048 ms) [F(1,38) = 54.09, MS, = 24,038, p < .001]; and participants took longer to respond to large problems (1209 ms) than to small problems (1068 ms) [F(1,38) = 69.99, MS, = 11,229, p < .001]. However, the main effects for task type and problem size were qualified by a Task Type × Problem Size interaction [F(1,38) = 28.06, MS, = 5,550, p < .001]: The problem size effect for production trials was larger (large - small: 1149 - 947 = 202 ms) than the problem size effect for verification trials (1268 - 1189 = 79 ms). Age did not interact with any factors (all ps > .50).

The verification RT data were analyzed separately so that the effect of response type could be examined. The 2 (age) × 2 (response type: true vs false) × 2 (problem size) analysis for these data showed main effects for age [F(1,38) = 10.55, MS, = 194,598, p < .01] (young = 1115 ms, older = 1342 ms), response type [F(1,38) = 57.60, MS, = 11,682, p < .001] (true = 1164 ms, false = 1293 ms), and problem size [F(1,38) = 24.61, MS, = 10,123, p < .001] (small = 1189 ms, larger = 1268 ms). The main effects for response type and problem size, however, were qualified by a Response Type × Problem Size interaction [F(1,38) = 9.76, MS, = 5,605, p < .01]: The problem size effect for “true” trials (large - small: 1222 - 1106 = 116 ms) was larger than that for “false” trials (1314 - 1272 = 42 ms). Again, age did not interact with any other factor.

To examine the effect of split in the verification task, the “false” trials were analyzed separately using a 2 (age) × 2 (split size: small vs large) × 2 (problem size) mixed ANOVA. Older adults took longer to respond (1421 ms) than young adults (1115 ms) [F(1,38) = 12.09, MS, = 208,477, p < .01], and participants took longer to respond to large-split trials (1329 ms) than to small-split trials (1263 ms) [F(1,38) = 10.36, MS, = 16,836, p < .01].

Error data. — The error data for the overall analysis were analyzed using a 2 (age) × 2 (task type) × 2 (problem size) ANOVA (see Table 1). This analysis found no age differences (p > .10). There were main effects for task type [F(1,38) = 33.23, MS, = .0025, p < .001] (mean percent error: verification = 7.0%, production = 11.5%) and problem size [F(1,38) = 5.91, MS, = .0035, p < .05]. However, these main effects were qualified by a Task Type × Problem Size interaction [F(1,38) = 8.18, MS, = .0028, p < .01]. When the verification error data were analyzed separately in order to compare true versus false trials, there was a main effect for response type [F(1,38) = 6.33, MS, = .014, p < .05] (mean percent error: true = 5.8%, false = 8.1%), and there was an Age × Problem Size interaction [F(1,38) = 6.61, MS, = .0535, p < .05]. Older adults showed a reverse problem size effect — their error rate on small problems (9.8%) was higher than that on large problems (7.0%) — whereas young adults showed a typical problem size effect (small = 4.5%, large = 6.6%). Nevertheless, there was no overall age difference for verification errors (p > .10). Finally, the verification error analysis for false trials only that examined split effects showed an Age × Problem Size interaction [F(1,38) = 6.00, MS, = .009, p < .05], and a Split Size × Problem Size interaction [F(1,38) = 11.92, MS, = .005, p < .05]. As was the case for the overall verification error analysis, the interaction of age and problem size resulted from the reverse problem size effect for older adults (large - small: 5.9 - 10.9 = -5.0%), while young adults showed a typical problem size effect (8.5 - 6.0 = 2.5%). Again, it is important to note that there were no overall age differences for the split analysis (p > .50). Finally, the Split Size × Problem Size interaction occurred because small problems resulted in more errors for large splits (small split = 7.6%, large split = 9.3%), whereas the reverse was true for large problems (small split = 10.2%, large split = 4.3%).

Because older adults showed a reverse problem size effect for errors, but not for latencies, an analysis of covariance (ANCOVA) was conducted using RT as the dependent variable and errors as the covariate. The key issue was whether a speed-accuracy tradeoff existed for older adults. The ANCOVA revealed no Age × Task or Age × Task × Problem Size interactions (p > .14). Nevertheless, it appears that for older adults, slopes of RT as a function of problem size were attenuated when errors were taken into consideration. This speed-accuracy tradeoff was not extreme enough to result in subadditivity on the part of older adults relative to young adults.

Regression analyses. — To obtain indices of central and peripheral processing, we performed separate regression analyses for individual participants using problem size as the predictor variable and RT as the criterion variable (after Geary & Wiley, 1991; we used the correct product to predict RT). The estimated slope is an index of retrieval efficiency, and the estimated intercept is assumed to reflect peripheral processing time (Cerella, 1991). The correlation, over participants, of slopes and intercepts was .03, and this is consistent with the notion that these coefficients reflect different processes. We collapsed across task type because the earlier ANOVAs revealed that age and task type did not interact. The average correlation of problem size and regression time was .45. ANOVA revealed no significant effect of age on slope (mean slope: older = 4.20; young = 3.37; p > .29), but the mean intercept of older adults (1137 ms) was significantly higher than that of young adults (958 ms) [F(1,38) = 10.77, MS, = 30,043, p < .01].

A separate regression analysis was conducted on each individual’s tie problems (e.g., 2 × 2) for the operands 2–9. Tie problems tend to show smaller problem size effects (Miller, Perlmutter, & Keating, 1984). Older adults exhibited steeper slopes (2.97) than young adults (0.64) [F(1,38) = 6.93, MS, = 7.82, p < .05]; however, the age groups did not differ in intercepts (older = 1135 ms, young = 1039 ms) (p > .13). Note that these results are exactly reversed from the overall analysis of the entire set of 64 problems (for which there was no age difference in slope, but there was an age difference in intercept), and that this was the result of young adults’ slope values dropping without a corresponding drop in older adults’ slope values. The discrepancy between the overall and the tie analyses suggested that for young adults the processing of tie problems differed from that of no-tie problems. Therefore, we
analyzed separately the no-tie problems (i.e., 52 problems — omitting the 8 ties and 4 problems with the same answer as a tie problem, e.g., $4 \times 4 = 16$ and $2 \times 8 = 16$). Problems with the same answer as the tie problems were also excluded because network models predict that these answers are interconnected in the memory network (Ashcraft, 1992). When slopes and intercepts were analyzed for an effect of age, the intercepts of older adults were significantly higher than those of young adults (older $= 1019$ ms, young $= 857$ ms) [$F(1,38) = 8.88$, $p < .01$], but there was no age effect for slopes (older $= 6.20$, young $= 5.75$, $p > .70$).

Hierarchical regression. — To determine if age accounted for a unique portion of experimental variance in Experiment 1, we predicted the intercept using the slope and age (in this order), and we predicted the slope using the intercept and age (in this order). The no-tie data were used for this analysis because responses to tie problems appeared to depend on special strategies for the young adults. When slope served as the criterion variable and intercept was entered before age, age was not a significant predictor ($p > .70$). However, when intercept served as the criterion variable and slope was entered as a predictor variable before age, age was a significant predictor [$F(1,37) = 8.77$, $p < .01$]. Note that this test of the significance of age when slope is entered first indicates that the increment in $R^2$ was significant for slope and age predicting the intercept ($R^2 = .17$) compared with just the slope predicting the intercept ($R^2 = .004$).

Brinley analysis. — A Brinley (1965) analysis was carried out using data from the 12 task conditions (4 pairs of means from each of the 3 different analyses: task type, response type, and split type; see Table 1). The mean latencies of the older adults were regressed on the mean latencies of the young adults. The best-fitting linear model for Experiment 1 was $Y = 1.27(X) - 78$ ($R^2 = .90$).

**DISCUSSION**

In Experiment 1, although older adults took longer than young adults to respond to both verification and production trials, older adults did not show significantly larger problem size or split effects. Also, task type did not interact with age and problem size. Response time was regressed on problem size to obtain parameters that could be interpreted as indicators of central and peripheral processing. Except for tie problems (e.g., $2 \times 2$), mean slope did not depend upon age. These results suggest that older adults retrieved basic multiplication facts at approximately the same rate as young adults (958 ms), that a manipulation designed to reveal potentially different retrieval processes (i.e., verification vs production) did not affect this situation. However, the mean regression intercept of older adults was significantly higher (1137 ms) than that of the young adults (958 ms). The intercept of the problem size regression equation reflects encoding (input) and response execution (output) processes (e.g., Ashcraft, 1992; Geary & Wiley, 1991). Experiment 2 was designed to examine further encoding as the source of age differences in peripheral processing in mental arithmetic.

The results from Experiment 1 provide evidence that process-specific slowing occurred. Older adults showed higher intercepts than young adults (a measure of peripheral processing speed), but there was no age effect on slopes (a measure of central processing speed). It is possible that this study lacked adequate power to detect an age difference in mean slope values. However, even if the small (but statistically nonsignificant) difference in mean slope observed in this experiment were a true effect, and we conducted a study with sufficient power to declare the effect significant (and it would take approximately 2,000 subjects to achieve a .80 power level with the .12 effect size in this experiment), we would still be faced with the result that the magnitude of the age effect on intercepts is vastly larger than that for slopes (see Frick, 1995). This result, too, would be consistent with the conclusion that age-related slowing is process-specific.

**Experiment 2**

In Experiment 2 all trials involved a multiplication production task, and exposure duration and problem size were varied. There is considerable evidence that decreasing exposure duration results in poorer stimulus quality (e.g., Allen, Madden, & Slane, 1995a; Allen et al., 1996; Loftus, Busey, & Senders, 1993). Evidence suggests that, for brief exposure durations, individuals are forced to use considerable error correction (or normalization) processing in order to overcome the degraded stimulus information (Allen & Emerson, 1991; Allen, Wallace, & LoSchiavo, 1994; Allen et al., 1995b). If the normalization processing is successful, responses should be correct, but slower than with longer exposure durations. If normalization processing is not successful, reducing exposure duration should increase errors. Thus, if age differences in overall mental arithmetic response times result from an age-related encoding deficit, then making encoding more difficult should exacerbate the age effect. Geary et al. (1993) proposed an encoding decrement model of age differences in mental arithmetic that attributed age effects in mental arithmetic performance to an age difference in encoding but not in central processes. The present Experiment 2 is a specific test of this model: If older adults require longer than young adults to respond to mental arithmetic problems due to encoding differences (i.e., a specific process difference), then there should be an Age $\times$ Exposure Duration interaction but no Age $\times$ Problem Size interaction. Alternatively, if age differences are generalized across processing stages, (a) then there should be slope and intercept effects for the analyses of the problem size regressions; (b) age should not account for unique variance in hierarchical regression; and (c) age differences should be accounted for with a single slowing function using Brinley analysis (Cerella, 1985).

**METHOD**

**Participants.** — A total of 48 subjects participated in Experiment 2. The 24 young adults (mean age = 24.1 years, range 18–32 years) were Cleveland State University undergraduates who participated for course credit in an introductory psychology course. The 24 older adults (mean age = 72.9 years, range 61–80) were Project 60 volunteers who were paid $15.00 for their participation. Participants were
all screened for near visual acuity of 20/40 using the Rosenbaum pocket vision screener.

All participants were tested on the WAIS–R Vocabulary and Digit Symbol Substitution Task subscales (Wechsler, 1981), and provided information concerning the number of years of education completed. There were no age differences for years of education (young = 14.4 years, older = 14.0 years). However, young adults scored significantly higher (71.2) than older adults (47.8) on the WAIS–R Digit Symbol Substitution Task \(F(1,46) = 81.09, p < .001\), whereas older adults scored higher (53.2) than young adults (46.9) on the WAIS–R Vocabulary subscale \(F(1,46) = 7.35, p < .01\).

**Materials.** — The apparatus used in Experiment 2 was the same as that used in Experiment 1.

**Procedure.** — For all trials, problems were presented horizontally in the center of the computer monitor. Each number subtended approximately 0.28° of horizontal visual angle and 0.56° of vertical visual angle. The average problem subtended approximately 3.30° of visual angle. When the participant stated an answer to the problem, the voice-activated response key stopped the clock that was initiated when the problem was first presented. The participant’s spoken answer for each trial was then entered into the computer manually by the experimenter.

Problem exposure duration was varied. Three different exposure durations were used: until the participant responded, 600 ms, and 300 ms. For the experimental trials exposure duration was blocked, and the order of exposure-duration blocks was counterbalanced across participants. There were 100 trials for each of the three exposure durations, for a total of 300 experimental trials per participant. Instructions gave equal emphasis to speed and accuracy. Participants received 20 practice trials prior to experimental trials (practice trials consisted of stimuli borrowed from the same set of 100 problems used in the experimental trials).

**RESULTS**

**Latency data.** — Correct response latencies between 100 ms and 3 sec were used in the RT analyses for Experiment 2 (correct trials outside of this interval were treated as guessing errors; these trials constituted only 0.9% of the total trials). Table 2 presents RT data as a function of age, exposure duration (presentation–until-response, or PUR; 600 ms, and 300 ms), and problem size for problems with operands varying from 2 to 9. The results of the 2 (age) \(\times\) 3 (exposure duration) \(\times\) 2 problem size: small vs large) ANOVA showed that older adults took somewhat (although not significantly) longer to respond (1116 ms) than young adults (1034 ms) \(F(1,46) = 3.01, MS_e = 160.958, p = .0896\). Participants, in general, took longer to respond to large problems (1169 ms) than to small problems (980 ms) \(F(1,46) = 89.30, MS_e = 28.780, p < .001\) and took longer to respond to more briefly presented problems (PUR = 1027 ms, 600 ms = 1016 ms, and 300 ms = 1183 ms) \(F(2,92) = 33.96, MS_e = 24.701, p < .001\).

However, there was an Age \(\times\) Exposure Duration interaction \(F(2,92) = 4.83, MS_e = 24.701, p < .05\). Younger adults showed a larger exposure duration cost (i.e., 300 ms – PUR) than older adults. For older adults the exposure duration cost was 1194 – 1106 = 88 ms, whereas for young adults the exposure duration cost was 1172 – 947 = 225 ms. Finally, there was a Problem Size \(\times\) Exposure Duration interaction \(F(2,92) = 6.65, MS_e = 6.403, p < .01\) that reflected the larger exposure duration cost for small problems (1106 – 908 = 198 ms) than for larger problems (1260 – 1144 = 116 ms).

**Error data.** — The error data for the overall analysis were analyzed using a 2 (age) \(\times\) 3 (exposure duration) \(\times\) 2 (problem size) ANOVA (see Table 2 for means). This analysis found no age differences \(p > .40\). Indeed, the only reliable effect was for problem size \(F(1,46) = 15.68, MS_e = .012, p < .001\) (mean percent error: large problems = 10.5%, small problems = 5.3%).

**Regression analyses.** — As was the case for the Experiment 1 data, we regressed the response times of each participant on problem size. Because exposure duration interacted with age, a separate regression equation was estimated for each participant for each exposure duration (a total of 144 separate regression equations, mean \(r = .59\)). As noted earlier, the slopes of these regression equations are

<table>
<thead>
<tr>
<th>Problem Size</th>
<th>PUR Mean</th>
<th>SD</th>
<th>600 ms Mean</th>
<th>SD</th>
<th>300 ms Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young RT</td>
<td>836</td>
<td>113</td>
<td>894</td>
<td>184</td>
<td>1089</td>
<td>131</td>
</tr>
<tr>
<td>Old RT</td>
<td>981</td>
<td>192</td>
<td>961</td>
<td>191</td>
<td>1122</td>
<td>163</td>
</tr>
<tr>
<td>Young % error</td>
<td>5.1</td>
<td></td>
<td>5.1</td>
<td></td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>Old % error</td>
<td>4.8</td>
<td></td>
<td>4.8</td>
<td></td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young RT</td>
<td>1057</td>
<td>274</td>
<td>1072</td>
<td>245</td>
<td>1255</td>
<td>206</td>
</tr>
<tr>
<td>Old RT</td>
<td>1231</td>
<td>245</td>
<td>1137</td>
<td>252</td>
<td>1264</td>
<td>197</td>
</tr>
<tr>
<td>Young % error</td>
<td>9.4</td>
<td></td>
<td>10.4</td>
<td></td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td>Old % error</td>
<td>11.9</td>
<td></td>
<td>10.6</td>
<td></td>
<td>11.5</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Reaction Time (RT, in Milliseconds) and Mean Percent Error Data for Problems From Experiment 2 for the Age Group (Young vs Older Adults) \(\times\) Problem Size (Small vs Large) \(\times\) Exposure Duration (Presentation Until Response (PUR), 600 ms, and 300 ms) Analysis**
assumed to index central processing speed (the larger the slope, the larger the problem size effect), and the intercepts are assumed to measure peripheral processes (e.g., encoding or response execution). The estimated slopes and intercepts were analyzed using separate mixed 2 (age group) × 3 (exposure duration: PUR, 600 ms, and 300 ms) ANOVAs. The slope ANOVA revealed neither an effect of age nor an interaction of age with exposure duration (slopes: presentation-until-response: older = 6.07, young = 6.88; 600 ms: older = 4.70, young = 5.22, 300 ms: older = 4.73, young = 4.31) (p > .50). The intercept ANOVA revealed significantly higher intercepts for older adults (982 ms) than for young adults (891 ms) [F(1,46) = 4.60, MS = 65,595, p < .05], and that older adults actually showed less increase in intercept than young adults as exposure duration decreased (intercepts: presentation-until-response: older = 949 ms, young = 761 ms; 600 ms: older = 933 ms, young = 848 ms; 300 ms: older = 1065 ms, young = 1063 ms). That is, the Age × Exposure Duration interaction was significant [F(2,92) = 5.00, MS = 20,746, p < .01].

Regression equations for individuals were also computed for the eight tie problems with operands between 2 and 9 (an additional 144 regression equations). The slope and intercept data were then analyzed using the same 2 × 3 ANOVA design used for the full 64 problems described in the preceding paragraph. The resulting slope analysis revealed no significant effects (p > .18) (older slope = 1.42, young slope = 2.58), although, once again, the tie slopes were much shallower than the overall slopes, suggesting that processing was somehow different for these trials. The intercept analysis for tie problems also failed to reveal any significant effects for age, although the Age × Exposure Duration interaction approached significance (p = .0733). As was the case for the analysis of intercepts for the full 64-problem set, the tie intercept analysis showed a trend in which older adults showed subadditivity (intercepts: presentation-until-response: older = 971 ms, young = 725 ms; 600 ms: older = 1002 ms, young = 887 ms; 300 ms: older = 1080 ms, young = 1060 ms).

Hierarchical regression. — Because the 600 ms and 300 ms conditions resulted in either correlated slopes and intercepts or unreliable slopes, only the estimated coefficients from the PUR no-tie problems were subjected to hierarchical regression analyses. When slope and age, in this order, were used to predict the intercept, age accounted for a statistically significant portion of variance [F(2,46) = 10.21, p < .05]. However, when the intercept and age, in this order, were used to predict the slope, then age was not a statistically significant predictor (p > .8). Consequently, the hierarchical regression results from PUR condition of Experiment 2 replicated the results from Experiment 1.

Brinley analysis. — The Experiment 2 latencies were used to conduct a Brinley analysis based on the 69 cell means (3 Exposure Durations × 23 Problem Sizes) collapsed across subjects. The best-fitting least-squares, two-parameter (i.e., one slope and one intercept) regression equation was \( Y = .73X + 369 \left( R^2 = .62 \right) \). According to Cerella (1985), such a linear effect suggests an age difference in peripheral processing — except the slope of less than one indicates that the age decrement was on the part of the young adults. We also tested whether a four-parameter model (one slope and a separate intercept for each exposure duration) would represent a significant improvement over the two-parameter model using the \( F \)-test suggested by Fisk, Fisher, and Rodgers (1992). The best-fitting least-squares four-parameter model (with one slope and three intercepts) was \( Y = .83X + 325 - 110(D_1) + 24(D_2) \left( R^2 = .68 \right) \), in which \( Y \) is predicted RT for older adults' problem size data, \( X \) is young adults' problem size data, and \( D_1 \) and \( D_2 \) are dummy variables that represent the 600 ms and 300 ms exposure durations, respectively. The \( F \)-test of the increment in variance accounted for by the four-parameter model compared with the two-parameter model was significant [\( F(2,67) = 6.52, p < .01 \)]. The finding that a four-parameter slowing function (one slope and three intercepts) accounted for significantly more variance than a two-parameter slowing function (one slope and one intercept) was consistent with the hypothesis of process-specific slowing.

DISCUSSION

Although age interacted with response time in this experiment, older adults showed smaller encoding effects for decreased exposure duration than did young adults. The shallower slope of the exposure duration effect for older adults occurred even though older adults had significantly higher intercepts. Older adults exhibited an overall encoding decrement, but this encoding decrement actually was least pronounced at the shortest exposure duration. Older adults showed intercept (an index of encoding speed) decrements relative to young adults at the PUR and marginally at the 600-ms exposure durations in the regression analysis done on individuals. However, in the most difficult condition (the 300-ms exposure duration), there were no effects of age on intercept. Furthermore, there were no age differences in mean RT or errors in this most difficult condition (see Table 2). There was no effect of age on slope, an index of retrieval speed, at any exposure duration.

The hierarchical regression analyses of slopes and intercepts from the PUR condition were also consistent with a process-specific decrement: When slope and age were used to predict intercepts, age contributed significantly to the prediction of the intercept, but when the intercepts and age were used to predict slopes, age was not a significant predictor. With the additional finding of fairly comparable reliabilities for slopes and intercepts (.54 and .67) and a very low correlation between slopes and intercepts (.14), the hierarchical regression analyses from Experiment 2, along with those from Experiment 1, make a strong argument for process-specific effects. In addition, in the Brinley analysis, a four-parameter slowing function (i.e., a single slope and separate intercepts for each exposure duration) accounted for significantly more variance than a two-parameter slowing function (i.e., one slope and one intercept). Hence, the hierarchical regression analyses and the ANOVA analyses indicated that the effect of age differed for slopes and intercepts, and the Brinley analyses revealed that multiple slowing functions accounted for significantly more variance than did a single slowing function. These results are inconsistent
with a general slowing model (e.g., Cerella, 1985, 1991; Lima, Hale, & Myerson, 1991; Myerson et al., 1990).

General Discussion

In the present experiments we examined age differences in basic arithmetic fact retrieval when task type was varied (Experiment 1: verification vs production) and when encoding difficulty was varied by manipulating stimulus quality (Experiment 2: PUR, 600-ms, and 300-ms exposure durations). The results of Experiment 1 suggested that older adults' longer overall latencies (compared with young adults) were the result of a peripheral process decrement, but not a fact retrieval (or central process) decrement. There were no age differences in the problem size effect or in the slope analysis (a measure of processing speed per unit of product size — typically assumed to be a measure of retrieval speed, e.g., Allen et al., 1992; Ashcraft, 1992; Geary & Wiley, 1991). However, there was an age decrement on the part of older adults in the intercept analysis (a measure of encoding and/or response speed). For both Experiments 1 and 2, the correlation between slopes and intercepts was very low (.03 and .14, respectively), which implies that these two dependent variables were measuring different processes. The finding that age did not interact with task type or with problem size for either RT or errors also is consistent with the proposal that even different types of retrieval do not result in age differences in central processes (i.e., verification and production may rely on somewhat different retrieval processes, Campbell & Tarling, 1996). Consequently, the results from Experiment 1 indicated that although there were age differences in the processing of mental arithmetic problems, there was no appreciable evidence of age differences for central processes. These conclusions are consistent with those of Geary and Wiley (1991), Allen et al. (1992), and Geary et al. (1993).

Experiment 2 further examined age differences in multiplication production encoding and retrieval by manipulating exposure duration. Exposure duration manipulations influence encoding efficiency by varying stimulus quality (e.g., Allen et al., 1994, 1995b; Loftus et al., 1993). Although older adults exhibited subadditivity for exposure duration, given that older adults, overall, showed longer latencies and higher intercepts than did young adults, the data from Experiment 2 show an age difference for encoding. Furthermore, as was the case in Experiment 1 (and for all other aging studies that have specifically manipulated fact retrieval), there were no age differences in arithmetic fact retrieval speed.

Hierarchical regression. — Key support for the conclusion that age has different effects on central and peripheral processes was the finding that age was a significant predictor of intercepts even when the effect of the slopes was extracted first — yet age was not a significant predictor of slopes when the intercept variable was entered into the regression equation before age. This was found for both Experiments 1 and 2. In each experiment the reliabilities of slopes and intercepts (Experiment 1: .60, .86; Experiment 2: .54, .67) were much higher than the correlation between slopes and intercepts (Experiment 1 = .03, .14). Therefore, interpreting the hierarchical regression analyses as suggested by Salthouse and Coon (1994), the present results support the hypothesis of process-specific age differences. The magnitude of the age-related decrement in peripheral processing is greater than the magnitude of the decrement in central processing (if there is such a decrement).

Implications of the present results for theories of aging. — The present mental arithmetic results revealed that localized age differences in peripheral processes occurred for both verification and production tasks (Experiment 1). These comparable age differences for problem size effects across task type suggest that there are no appreciable age differences in retrieval for these tasks. Furthermore, the finding in Experiment 2 that a four-parameter slowing function (one slope and three intercepts) accounted for significantly more variance than a two-parameter (one slope and one intercept) slowing function suggested that encoding (as measured by intercepts) was sensitive to age but that central processes (as measured by slopes) were not. Finally, both experiments showed through hierarchical regression that age differences could be predicted by intercept data but not by slope data. Earlier studies suggested the existence of these process-specific age differences (e.g., Allen et al., 1992; Geary et al., 1993; Geary & Wiley, 1991; Rogers & Fisk, 1991), but the present methods allowed a more explicit test of this hypothesis.

Skill level and semantic memory. — The present results for highly overlearned mental arithmetic facts (also see Allen et al., 1992; Geary et al., 1993; Geary & Wiley, 1991), the results from visual word recognition studies (e.g., Allen, Madden, & Crozier, 1991; Allen et al., 1993, 1995a; Cerella & Fozard, 1984; Madden, Pierce, & Allen, 1993; Laver & Burke, 1993), typing studies (e.g., Salthouse, 1984), and bridge playing (e.g., Charness, 1987) all indicate that age differences vary across processing stages. It is important to note that all of these studies are based on tasks for which participants are either highly skilled in a motor task (i.e., a procedural memory task) or a lexical access, mental arithmetic, or bid retrieval task (i.e., semantic memory tasks). For these sorts of tasks older adults tend to show peripheral-process (input and output) decrements but not central-process (e.g., retrieval and decision stage) decrements (Allen et al., 1992; Charness, 1987). These results are almost the exact opposite to those found by Cerella (1985, 1991) and by Myerson et al. (1990). In the meta-analyses of Cerella (1985), older adults showed larger central-process decrements than peripheral-process decrements. Interestingly, though, all of the studies analyzed by Cerella were either episodic memory or visual search experiments (this was also the case for the Myerson et al., 1990, meta-analysis). We suggest that both views may have substantial support in accounting for certain data. For the present results it could have been the case that high levels of early exposure to arithmetic on the part of older adults prevented age-related declines in lexical tasks — at least when compared to younger adults with relatively less exposure to arithmetic (Geary et al., 1996).
REFERENCES


Received March 21, 1996
Accepted October 16, 1996