# Variability in Reaction Time Performance of Younger and Older Adults

David F. Hultsch, Stuart W. S. MacDonald, and Roger A. Dixon

Department of Psychology, University of Victoria, British Columbia, Canada.

**Age differences in three basic types of variability were examined: variability between persons (diversity), variability within persons across tasks (dispersion), and variability within persons across time (inconsistency). Measures of variability were based on latency performance from four measures of reaction time (RT) performed by a total of 99 younger adults (ages 17–36 years) and 763 older adults (ages 54–94 years). Results indicated that all three types of variability were greater in older compared with younger participants even when group differences in speed were statistically controlled. Quantile-quantile plots showed age and task differences in the shape of the inconsistency distributions. Measures of within-person variability (dispersion and inconsistency) were positively correlated. Individual differences in RT inconsistency correlated negatively with level of performance on measures of perceptual speed, working memory, episodic memory, and crystallized abilities. Partial set correlation analyses indicated that inconsistency predicted cognitive performance independent of level of performance. The results indicate that variability of performance is an important indicator of cognitive functioning and aging.**

RESEARCHERS examining cognitive functioning in adulthood have primarily been interested in age-related differences or changes in level of performance. Methodologically, this emphasis has translated into comparisons of average performance across different age groups (i.e., cross-sectional designs) or examination of changes in average performance within persons across time (i.e., longitudinal designs). Research on average age-related differences and changes in cognition has been useful, but it has reflected certain assumptions about the nature of human development. Specifically, this emphasis is rooted in the assumption that either the behaviors of interest are stable over time or that the trajectory of change that does occur is similar for all persons. This assumption with respect to level of performance represents one instantiation of a more general stability perspective that has dominated developmental research (Gergen, 1977; Nesselroade & Featherman, 1997). As noted by Nesselroade and Boker (1994), however, the concepts of stability and variability are logically dependent on one another—defining one demands consideration of the other.

Before we consider substantive questions related to variability of cognitive performance and aging, it is important to define selected terms. There are multiple classifications of types of stability and variability (e.g., Alwin, 1994), and sometimes the same label has been applied to different types (e.g., Christensen, Mackinnon, Korten, Jorm, Henderson, & Jacomb, 1999; Shammi, Bosman, & Stuss, 1998). We define three different types of variability by considering the minimum conditions necessary to observe it in relation to persons, measures, and occasions (Cattell, 1966; Nesselroade & Ford, 1985). First, one can consider differences between persons measured on a single task on a single occasion. Such variability between persons is typically referred to as interindividual differences, or *diversity*. Second, one can examine variability associated with measuring a single person once on multiple tasks (see Appendix, Note 1). In this case, the variability is in the profile of relative performance across measures, sometimes referred to as intraindividual differences, or *dispersion*. The third type of variability is defined by the minimum condition of measuring a single person on a single task at multiple occasions. Variability in performance across occasions has been labeled intraindividual variability (Li, Aggen, Nesselroade, & Baltes, 2001), or *inconsistency* (Shammi et al., 1998). The latter two types of variability refer to variability within persons.

There is evidence that aging is associated with increases in all three types of variability on cognitive tasks, although there are caveats to this assertion. By far, the largest amount of data available is relevant to diversity. Reviews of the literature have pointed to increasing interindividual differences in cognitive performance with increasing age. For example, Nelson and Dannefer (1992) reported that 79% of the studies on cognition and aging they reviewed showed increases in variability with age. In a more formal meta-analysis, Morse (1993) examined age differences in the coefficient of variability for measures of reaction time (RT), memory, and intelligence. Measures of RT, memory, and fluid abilities showed increasing diversity with age, whereas measures of crystallized intelligence did not. Similar findings have been reported by Christensen and associates (1994) for a large probability sample of older adults. In addition, recent longitudinal studies have found diverging patterns of cognitive change in adulthood (Christensen, Mackinnon, Korten, Jorm, Henderson, Jacomb, & Rodgers, 1999; Hultsch, Hertzog, Dixon, & Small, 1998; Rabbitt, 1993; Schaie, 1996). For example, Hultsch and colleagues found significant increases in variability over 6 years for seven of nine cognitive variables. In effect, individuals were becoming less alike as a function of individual differences in change.

The picture is considerably less clear when one considers variability within persons. Very few studies have examined dispersion of cognitive functioning in adulthood, and the pattern of results is not well established. Lindenberger and

*HULTSCH ET AL.*

Baltes (1997) examined intraindividual standard deviations for a set of 14 cognitive measures in a cross-sectional sample of adults ranging from 70 to 103 years of age. They found that variability among tasks did not differ with age for higher ability adults and actually decreased with age for lower ability adults. In contrast, Christensen, Mackinnon, Korten, Jorm, Henderson, and Jacomb (1999) reported just the opposite. They found increased dispersion of scores across measures of speed, memory, and spatial functioning in their cross-sectional sample of adults (70 to 90 years). They did not, however, observe any increase in dispersion over a longitudinal interval of 3.5 years. Dispersion has also been assessed by examining the stability of crystallized intelligence relative to other cognitive domains (e.g., Rabbitt, 1993). In this approach, a score indicating the deviation of a particular ability from an indicator of crystallized intelligence (usually a measure of vocabulary) is computed for each person. Increasing deviation scores presumably suggest greater dispersion of abilities. Two such studies have reported finding increases in dispersion with increasing age (Christensen, Mackinnon, Korten, Jorm, Henderson, & Jacomb, 1999; Rabbitt, 1993).

In the case of inconsistency, greater intraindividual variability has been observed for older adults compared with younger adults, at least for some tasks. Several studies have shown that inconsistency across trials on RT tasks increases with age (Anstey, 1999; Fozard, Vercruyssen, Reynolds, Hancock, & Quilter, 1994; Salthouse, 1993), although some researchers have suggested this increase can be accounted for by individual differences in mean-level performance (e.g., Salthouse, 1993; Shammi et al., 1998). In a particularly interesting approach, Ratcliff (1979) and others have shown that it is possible to fit explicit mathematical functions to empirical response time distributions across a wide range of tasks and conditions. This approach yields several parameter estimates of the response time distribution, including its variability and skew. Although this approach has been used largely with young adult samples, studies contrasting younger and older adults have all found increased inconsistency in response time distributions with increasing age (Spieler, Balota, & Faust, 1996; West & Baylis, 1998; West, Murphy, Armilio, Craik, & Stuss, in press).

In addition to inconsistency across trials within a session, intraindividual variability may also be observed across multiple testing occasions. For example, Hertzog, Dixon, and Hultsch (1992) examined cross-occasion inconsistency in story recall by testing seven older women for up to 2 years. They found substantial intraindividual variability in performance across occasions, and more than 20% of this variability was reliable variance that was not associated with practice, different stories, or other systematic changes over time. Similarly, Rabbitt, Osman, and Moore (2001) measured both within-session and across-session inconsistency in older adults' RT on a letter identification task. They found that greater intraindividual variability was associated with poorer performance on the Culture Fair Intelligence Test for both trial-to-trial and week-to-week intervals. In terms of age differences, Li and colleagues (2001) examined intraindividual variability for a set of memory and sensorimotor variables across 13 biweekly sessions in a sample of 24 older adults age 64 to 86 years. They found that variability was positively correlated with age for most sensorimotor measures and one of the memory measures.

There are both practical and theoretical implications associated with the possibility of age-related increases in inconsistency. From a clinical perspective, it suggests that one occasion of measurement may not provide an adequate assessment of cognitive competence (Dixon, Hertzog, Friesen, & Hultsch, 1993; Stuss, Pogue, Buckle, & Bondar, 1994). Intraindividual variability in performance may be particularly significant in the assessment of individuals whose disorders are mild or not easily definable (Gordon & Carson, 1990; Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000; Stuss et al., 1994). Similarly, Rowe and Kahn (1997) have suggested that intraindividual variability may be a risk factor predictive of successful aging.

From a theoretical perspective, examination of intraindividual variability may provide insight into the operation of cognitive systems. On the one hand, measurement of intraindividual variability and other characteristics of response distributions may provide information reflective of the operation of different cognitive processes (although it is unlikely that specific parameters will be influenced solely by specific processes). For example, Spieler, Balota, and Faust (2000) demonstrated that spatial and attribute selection processes influenced separate parameters of the RT distribution. Similarly, Hockley (1984) experimentally separated distribution parameters for four different cognitive tasks. On the other hand, a number of theorists have suggested that inconsistency in performance may be an indicator of neurological disturbance (e.g., Hendrickson, 1982; Jensen, 1982; Li & Lindenberger, 1999). For example, Li and Lindenberger used computational simulations to demonstrate that increasing random variability of the networks led to decreases in the level of performance and simultaneously to increases in the magnitude of between-network variability and strength of cross-task intercorrelations. These results are consistent with the view that both decreases in level of performance and increases in interindividual variability with age might be produced by greater intraindividual variability in neurobiological mechanisms (Myerson, Hale, Wagstaff, Poon, & Smith, 1990; Welford, 1980).

The purpose of the present study was to examine age differences in all three types of variability identified previously—diversity, dispersion, and inconsistency—with particular attention to inconsistency (intraindividual variability). Data from two relatively simple and two relatively complex RT tasks were available for a sample of younger adults and a large sample of older adults spanning a 40-year age range. To our knowledge, there are no studies that have examined age differences in all three of the defined types of variability simultaneously.

We focused on three principal questions. First, we examined whether there were age differences in variability that were independent of age group differences in processing speed. Some investigators have suggested that between- and within-person differences in variability may simply be a function of age-related differences in slowing (Hale, Myerson, Smith, & Poon, 1988; Salthouse, 1993). It is also critical to dissociate systematic within-person changes (e.g.,

practice effects) from changes that reflect inconsistency in performance. Previous studies have not always addressed these issues. We hypothesized that all three types of variability would be greater for older than for younger adults, even when group differences in speed and systematic changes in performance were controlled for.

Second, we examined the relationships among various indicators of within-person variability. If intraindividual variability in RT is a function of relatively endogenous influences associated with deterioration of neurobiological mechanisms, then we would expect to observe relatively stable individual differences in such variability. That is, we would expect inconsistency on one RT task to correlate positively with inconsistency on other RT tasks. We were also interested in examining whether within-person variability across trials (inconsistency) was associated with within-person variability across tasks (dispersion).

Finally, we examined whether measures of intraindividual variability in RT are predictive of level of performance on other cognitive tasks. For example, if individual differences in inconsistency are indicative of central nervous system integrity, then we would expect to observe negative relationships between inconsistency in RT and level of performance on other cognitive tasks. We might also expect that such correlations would be higher for tasks that are more reflective of basic information-processing capacity than for tasks that are more influenced by acquired knowledge or skill. A critical question, however, is whether information about the inconsistency of an individual's responses tells us anything that is unique. Thus, we also examined whether individual differences in variability and level of RT performance are independent predictors of performance on other cognitive tasks.

## Method

This article is based on cross-sectional data from the Victoria Longitudinal Study (VLS). The design of the VLS consists of longitudinal sequences in which multiple cross-sectional samples of community-dwelling older adults (initially age 54–87 years) are retested at intervals of 3 years with new samples added at intervals of 6 years. Young adult (17–36 years) comparison samples are also tested every 6 years (at longitudinal Wave 1), but are not followed longitudinally. The general design, participants, measures, and procedures of the VLS have been described extensively elsewhere (see Dixon et al., in press; Hultsch et al., 1998), and therefore only unique and pertinent components of the method are summarized here.

### PARTICIPANTS

Data from 862 participants (546 women, 316 men) from Wave 3 of Sample 1 and Wave 1 of Sample 2 were used in the present analyses. Participants were divided into four age groups. The young (Y) group ($n = 99$; 54 women, 45 men) ranged from 17 to 36 years ($M = 23.17$, $SD = 4.97$); the young-old (YO) group ($n = 178$; 119 women, 59 men) ranged from 54 to 64 years ($M = 60.38$, $SD = 2.95$); the mid-old (MO) group ($n = 361$; 230 women, 131 men) ranged from 65 to 74 years ($M = 69.56$, $SD = 2.78$), and fi-

nally, the old-old (OO) group ($n = 224$; 143 women, 81 men) ranged from 75 to 94 years ($M = 79.33$, $SD = 3.72$).

The participants in the VLS exhibit the typical selectivity of longitudinal samples compared with the general population. The average education of the sample was 14.72 years ($SD = 3.03$), although there were differences among groups as a function of age, $F(3, 850) = 5.51$, $p < .001$, $\eta^2 = .02$, and gender, $F(1, 850) = 22.84$, $p < .001$, $\eta^2 = .03$. Although all groups were well educated, the Y group ($M = 15.20$, $SD = 2.44$) and the YO group ($M = 15.19$, $SD = 3.17$) had significantly more education than the OO group ($M = 14.14$, $SD = 3.16$). The MO group ($M = 14.70$, $SD = 2.97$) did not differ significantly from any other group. Men ($M = 15.37$, $SD = 3.10$) had significantly more education than women ($M = 14.33$, $SD = 2.92$).

Performance on a 54-item recognition vocabulary test (adapted from Ekstrom, French, Harman, & Dermen, 1976) indicated a high level of verbal ability ($M = 42.82$, $SD = 7.85$), although there was a significant age effect, $F(3, 850) = 78.36$, $p < .001$, $\eta^2 = .22$. As expected, the Y group ($M = 32.83$, $SD = 7.64$) showed lower vocabulary performance than the other age groups (YO: $M = 42.84$, $SD = 7.76$; MO: $M = 44.71$, $SD = 6.28$; OO: $M = 44.20$, $SD = 6.98$), which did not differ.

Self-reported health was evaluated in several ways, including a single-item rating of health relative to others and a questionnaire assessing the presence of 26 specific health conditions. More than 90% of participants rated their health as very good or good ($M = 0.63$, $SD = 0.71$, on a 5-point scale ranging from 0 = very good to 4 = poor), and there were no significant differences among the groups. As expected, there were age, $F(3, 850) = 17.11$, $p < .001$, $\eta^2 = .06$, and gender, $F(2, 850) = 23.77$, $p < .001$, $\eta^2 = .03$, differences in reported chronic conditions, as well as a significant interaction, $F(3, 850) = 4.19$, $p < .01$, $\eta^2 = .02$. Women and older adults reported more chronic conditions, although the differences between the genders decreased with increasing age: $Y_{men}$: $M = 0.69$, $Y_{women}$: $M = 2.50$; $YO_{men}$: $M = 1.83$, $YO_{women}$: $M = 2.52$; $MO_{men}$: $M = 2.46$, $MO_{women}$: $M = 2.67$; $OO_{men}$: $M = 3.06$, $OO_{women}$: $M = 3.41$.

### MEASURES AND PROCEDURE

The VLS measurement battery consists of multiple questionnaires, tests, and tasks focused on both cognitive and noncognitive variables. The test battery was administered during four testing sessions scheduled over a period of about 4–6 weeks. Tasks were administered in the same order to all participants. Data were collected from both samples within the same time frame (1992–1993).

### RT Tasks

The principal measures of interest were four multitrial computer-based RT tasks. Two of the measures assessed speed of responding to relatively simple nonverbal signals, whereas two of the measures involved speed of responding to more complex language-based stimuli. For all tasks, stimuli were presented on a computer monitor interfaced with a 386 IBM-compatible computer that controlled stimulus presentation and timing. Participants responded to stimuli by pressing keys on a custom-designed response con-

sole. Responses were recorded at an accuracy of plus or minus 1 ms.

*Simple reaction time (SRT).*—In the SRT task, participants were presented with a warning stimulus (***) followed by a signal stimulus (+) in the middle of the screen. Participants were instructed to press a key with their preferred hand as quickly as possible when the signal stimulus appeared. A total of 50 test trials were administered with 10 randomly arranged trials presented at each of five intervals separating the warning and signal stimuli (500, 625, 750, 875, and 1,000 ms). The measures used were the latencies of the 50 test trials.

*Choice reaction time (CRT).*—For CRT, a 3 × 3 grid matching the arrangement of keys on the response console was displayed on the screen. This array was used to instrument two-, four-, and eight-choice RT trials. The center square, corresponding to the center key in the response keypad, served as the home key for the participant's right forefinger. Each block of 10 trials required the participant to attend to two, four, or eight squares. A warning stimulus was presented, followed (after a delay of 1,000 ms) by the appropriate two-, four-, or eight-square matrix. One square contained an *O* and all the others contained *X*s. The participant's task was to press the key corresponding to the location of the *O*. Twenty trials were administered at each level of choice. The measures used were the latencies of all trials averaged across conditions.

*Lexical decision.*—In the lexical decision task, participants were presented with a string of five to seven letters on the computer screen and were asked to indicate as quickly as possible whether they formed an English word (e.g., *island* vs *nabion*). A total of 60 test trials were presented (30 words and 30 nonwords). The measures used consisted of the latencies of the 60 trials.

*Semantic decision.*—In the semantic decision task, participants were asked to judge as rapidly as possible the plausibility of sentences presented on the computer screen (e.g., *The tree fell to the ground with a loud crash* vs *The pig gave birth to a litter of kittens this morning*). A total of 50 sentences were presented, and latencies of the trials were used as the measures.

### Other Measures

In addition to the four latency-based tasks, indicators of perceptual speed, working memory, episodic memory, and crystallized abilities were examined as correlates of within-person variability. These domains range along a rough continuum from measures of basic processing resources to indicators of acquired knowledge. Inclusion of measures of each domain were based on previous confirmatory factor analyses (Hultsch et al., 1998). For each domain, a linear composite was created by standardizing and averaging the individual scores.

*Perceptual speed.*—This variable was defined by three paper-and-pencil measures. Two tasks from the Kit of Fac-

tor-Referenced Cognitive Tests (Ekstrom et al., 1976) required participants to make simple perceptual comparisons as rapidly as possible within a limited time period. In Identical Pictures, participants chose which one of five line drawings matched a target figure. In Number Comparison, participants indicated whether two strings of digits were identical or not. The third measure was the revised Wechsler Adult Intelligence Scale Digit Symbol Substitution task. Participants were given 90 s to transcribe as many symbols as possible into empty boxes on the basis of the digit–symbol associations specified in a coding key. For all three tasks, the measures consisted of the number of correctly completed items.

*Working memory.*—This domain was indexed by two widely used working memory tasks developed by Salthouse and Babcock (1991). Both tasks require storage of information and simultaneous processing of that information. In Computation Span, participants solved arithmetic problems while holding one number from each problem in memory for later recall. In Listening Span, participants listened to orally presented sentences and wrote answers to simple questions about each sentence while retaining the last word of each sentence for later recall. In each task, the number of items (problems, sentences) increased from one to seven, with three trials at each series length. For each task, the score used was the highest span (one to seven) correctly recalled on two out of three trials.

*Episodic memory.*—Both word and story recall tasks were used. Word recall consisted of immediate free recall of two lists of 30 English words selected from the total set of six lists (Hultsch, Hertzog, & Dixon, 1990). Each list consisted of 6 words from each of five taxonomic categories (e.g., birds, flowers) typed on a single page in unblocked order. Participants were given 2 min to study each list and 5 min to write their recall. The number of correctly recalled words from each of the two lists were used as the measures. Story recall was measured by immediate gist recall of two narrative stories about an event in the life (or lives) of an older adult (or couple). The total set of six stories was selected from a larger set of 25 structurally equivalent texts developed by Dixon, Hultsch, and Hertzog (1989). Each story was approximately 300 words and 160 propositions long. The stories were presented in typed booklets for study followed by written recall. Participants were given 4 min to read each story and 10 min to write their recall. Recall protocols were scored for gist recall using criteria described in Dixon and associates (1989). Reliability estimates of the scoring system across all possible pairs of scorers exceeded 90%. The total number of gist propositions recalled from each of the two stories were used as the measures.

*Crystallized abilities.*—Measures of world knowledge and vocabulary were used to index crystallized abilities. World knowledge was measured by two sets of 40 questions that tested individuals' recall of facts about multiple domains including science, history, literature, sports, geography, and entertainment (Nelson & Narens, 1980). The questions were presented in booklets, and participants wrote

their answers under self-paced timing conditions. The vocabulary measure consisted of performance on a 54-item multiple-choice (recognition) vocabulary test composed by concatenating three 18-item tests from Ekstrom and colleagues (1976). The number of correct responses on each task was used as the measure.

## DATA PREPARATION

We first examined the distributions of raw latency scores for outliers. Extremely fast or slow responses might reflect various types of errors (accidental key press, interruption of the task). To address these potential concerns, outlier scores were trimmed as follows. A lower bound for legitimate responses was set for each task on the basis of minimal response times suggested by prior research, and scores below this limit were dropped. The limits were SRT, 150 ms; CRT, 150 ms; lexical, 400 ms; and semantic, 1,000 ms. The upper bound was established by computing the mean and standard deviation separately for each of the age groups and dropping any trials exceeding the mean by three or more standard deviations. The number of trials dropped across the entire Persons × Trials data matrix was relatively small given the number of data points involved (SRT = 2.2%; CRT = 2.1%; lexical = 3.6%; semantic = 7.0%). Percentage of missing trials did not vary systematically across age groups. To avoid statistical problems associated with missing data, we then imputed values for the outlier trials by using a regression procedure in which missing value estimates were based on the relationships among responses across trials. Missing values were imputed using data from all individuals and trials available. Because dropping outlier scores and imputing the resulting missing values reduces variability, these data preparation strategies represent a conservative approach to examining the phenomenon.

## Results

The results are presented in five main parts. In the first three sections, we examine age differences in diversity, dispersion, and inconsistency in latency performance on the four main RT tasks. In the fourth section, we report correlations to examine relationships among the various measures of within-person variability. Finally, we examine the relationships of level of RT performance and intraindividual variability in RT performance to performance on measures of perceptual speed, working memory, episodic memory, and crystalized abilities. In particular, we used set correlation approaches to determine whether level and variability in RT performance are independent predictors of performance on other cognitive tasks.

### VARIABILITY BETWEEN PERSONS

We began by examining the question of whether there is increasing diversity in RT latency performance with increasing age. Table 1 shows the standard deviations for the four tasks as a function of age. The last column of the table reports Levene's test for homogeneity of variance, indicating there were significant group differences in variability for all four measures. However, there were also age differences in average latency as indicated by the means in Table

Table 1. *T*-Score Standard Deviations and Means of Reaction Time Performance for Four Tasks by Age Group

| Task | Age Group | | | | Levene's Test, $F(3, 858)$ |
| --- | --- | --- | --- | --- | --- |
| | Young | Young-Old | Mid-Old | Old-Old | |
| Simple reaction time | | | | | |
| *SD* | 3.32 | 4.86 | 5.75 | 6.78 | 14.72 |
| *M* | 44.19 | 48.32 | 50.48 | 53.12 | |
| Choice reaction time | | | | | |
| *SD* | 3.84 | 4.42 | 4.98 | 6.25 | 10.26 |
| *M* | 39.86 | 47.58 | 51.05 | 54.71 | |
| Lexical | | | | | |
| *SD* | 4.95 | 5.22 | 6.01 | 7.21 | 11.41 |
| *M* | 46.21 | 48.14 | 49.89 | 53.33 | |
| Semantic | | | | | |
| *SD* | 5.69 | 5.01 | 5.91 | 7.56 | 9.22 |
| *M* | 47.65 | 47.61 | 49.31 | 54.05 | |

*Note*: $p$s < .001.

1 (there were significant age differences in mean latency on all tasks, but because our focus is on variability, we do not report the results of the analyses on level of performance). Thus, group differences in variability may be an artifact of group differences in mean performance because larger standard deviations tend to be associated with larger means (Hale et al., 1988).

We addressed this issue using a regression approach previously implemented by Christensen and colleagues (1994). Specifically, we regressed each of the four RT tasks on age, yielding residual scores. Residuals were calculated with only the significant polynomial trends included in the analysis (both the linear and quadratic age trends were significant in all cases). Figure 1 shows the absolute value of the residual scores as a function of age for the four tasks. To examine diversity across age, the residual scores for each task were then regressed on the linear and quadratic age trends. Table 2 reports the results of these analyses. There was a significant linear trend for all tasks, indicating that performance was increasingly diverse with increasing age. Similarly, there was a significant quadratic trend for all tasks except SRT. For the CRT, lexical, and semantic tasks, the positive slopes indicate this trend is a function of an increasing rate of diversity in the oldest participants. In general, the magnitude of these significant trends was modest—typically less than 2% of the variance.

The previous analysis was not based on a continuous age range, and therefore it is possible that the significant age trends are largely a function of differences in diversity between the younger and older groups. To examine this issue, we repeated the analysis using only the older participants in the continuous age range of 54 to 94 years. Increasing diversity in performance with increasing age was again observed, but with one notable difference. Although the linear trends were again significant for all tasks, the quadratic trend was significant only for the lexical task. This suggests that the previously significant quadratic effects were largely a function of the contrast between the younger and older participants.

We previously noted that there were small but significant differences among the age groups in education. Although
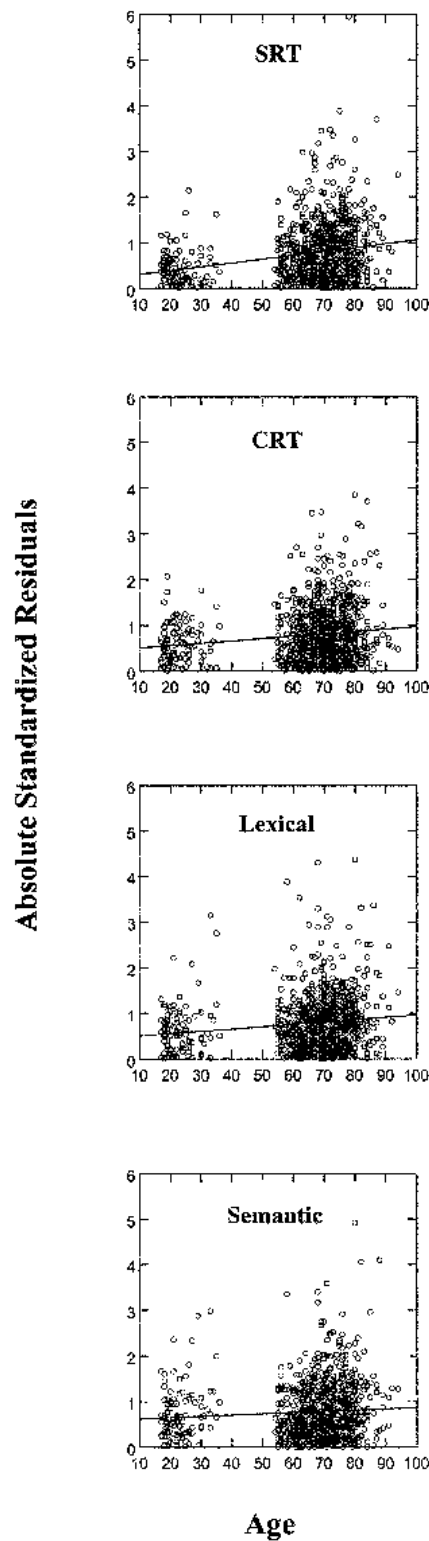
Figure 1. Scatter plots and regression lines for absolute standardized residual scores as a function of age for four reaction time (RT) tasks. SRT = simple reaction time; CRT = choice reaction time.

Table 2. Summary Table for Regression of Residuals on Linear and Quadratic Age Trends (All Ages)

| Predictor | β | R | $\Delta R^2$ | N |
|---|---|---|---|---|
| Simple reaction time | | | | |
| Linear age | .219 | .219 | .048** | 862 |
| Quadratic age | .251 | .224 | .002 | |
| Choice reaction time | | | | |
| Linear age | .146 | .146 | .021** | 862 |
| Quadratic age | .401 | .164 | .006* | |
| Lexical | | | | |
| Linear age | .137 | .137 | .019** | 862 |
| Quadratic age | .555 | .172 | .011** | |
| Semantic | | | | |
| Linear age | .076 | .076 | .006* | 862 |
| Quadratic age | .546 | .127 | .010** | |

*$p < .05$; **$p < .01$.

education was not strongly correlated with RT performance or variability, we reran the analyses partialing out education. The pattern of results remained the same.

## VARIABILITY WITHIN PERSONS

There are multiple indices that may be computed to examine intraindividual variability (Slifkin & Newell, 1998). Perhaps the simplest of these is the intraindividual standard deviation (ISD). An ISD can be computed across tasks to examine dispersion or across time (trials or occasions) to examine inconsistency. However, simply computing ISDs on raw scores is problematic. For most cognitive measures, one typically observes significant group differences in average level of performance. In addition, systematic changes over time (trials, occasions) associated with practice, different materials, and so forth may be present. These group and systematic time-related effects represent potential confounds for the analysis of intraindividual variability. For example, evidence of greater interindividual variability in older adults as indicated by an ISD computed on raw scores may simply reflect the fact that older adults are on average slower than younger adults. To address these issues, we partialed out the effects associated with age group, gender, trial, and all their interactions from the data before computing ISDs. This procedure produced residual scores that were uncontaminated by group differences in speed or accuracy of performance and systematic variation due to influences such as practice. These purified scores were then converted to $T$ scores to permit comparison of the tasks in the same metric. Figure 2 shows an example of the residual $T$ scores by trials for one task (SRT latency) for each individual participant graphed separately by age group. This figure shows that even though all systematic effects have been partialed out from the data (all groups have a mean of 50 and a standard deviation of 10), substantial individual differences in intraindividual variability remain (and appear to vary across group).

### Dispersion

To examine intraindividual variability across tasks, we computed ISDs for each individual over the purified residual $T$ scores of the four RT tasks. Lower values on this dis-
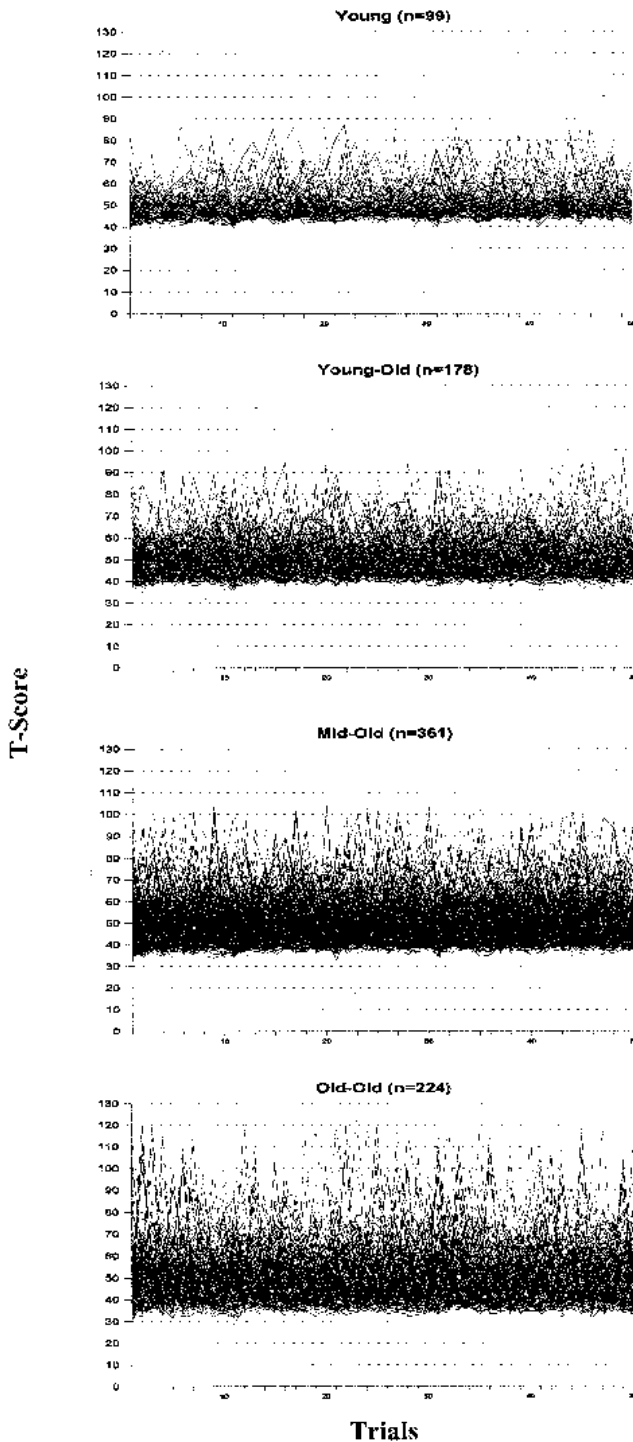
Figure 2. Simple reaction time residual latency *T* scores by trial (purified for age, gender, and trial effects) for each participant graphed separately by age group.

person score reflect relatively flat intraindividual profiles of performance across tasks, whereas higher values refer to relatively uneven profiles of performance.

A 4 (age) $\times$ 2 (gender) analysis of variance (ANOVA) revealed significant effects associated with age, $F(3, 854) = 11.82, p < .001, \eta^2 = .04$. Gender and the interaction of age

and gender were not significant. Average across-task ISDs increased as a function of age group (Y: $M = 3.91$, YO: $M = 4.15$, MO: $M = 4.76$, OO: $M = 5.33$), indicating there was increasing dispersion with increasing age. Post hoc analyses conducted using Tukey's honestly significant difference (HSD; $p < .05$) indicated that the OO group showed greater intraindividual variability across tasks than all other age groups. Similarly, the MO age group exhibited more dispersion than the YO or Y groups. The Y and YO groups, however, did not differ significantly. To estimate the magnitude of the group differences, we computed effect sizes using Cohen's $d$ for the comparisons where significant differences were observed (see Appendix, Note 2). Effect sizes ranged from small to medium according to Cohen's convention (small = .20, medium = .50, large = .80): OO/MO = 0.24, OO/YO = 0.53, OO/Y = 0.60, MO/YO = 0.28, MO/Y = 0.38. The results remained unchanged when education was covaried.

*Inconsistency*

To examine intraindividual variability over time, we computed ISDs for each individual over the purified residual trial scores separately for each of the four RT tasks. Higher scores on this measure indicate relatively inconsistent performance across trials, whereas lower scores indicate relatively consistent performance. We also computed the coefficient of variation in which each individual's ISD on any given task is divided by his or her average score on that task. This yields a measure of inconsistency relative to the individual's overall level of performance. Finally, we performed analyses to examine the characteristics of the ISD distributions.

*Mean ISDs.*—A 4 (age) $\times$ 2 (gender) multivariate analysis of variance computed on the ISD scores for the four tasks revealed significant omnibus effects associated with age, Wilks's $\lambda = .642$, $F(12, 2252) = 34.27, p < .001, \eta^2 = .14$, and gender, Wilks's $\lambda = .978$, $F(4, 851) = 4.69, p < .001, \eta^2 = .02$. The interaction was not significant. Figure 3 shows the mean ISDs on the four tasks by age group. Univariate ANOVAs indicated there were significant age group differences on all of the tasks: SRT: $F(3, 854) = 70.27, p < .001, \eta^2 = .20$; CRT: $F(3, 854) = 57.36, p < .001, \eta^2 = .17$; lexical: $F(3, 854) = 34.40, p < .001, \eta^2 = .11$; and semantic: $F(3, 854) = 43.20, p < .001, \eta^2 = .13$. We used Tukey's HSD to specify the age group differences for each task. For SRT, intraindividual variability increased significantly across each age group. That is, the OO group was more inconsistent than the MO group, the MO was more inconsistent than the YO group, and the YO group was more inconsistent than the Y group. The magnitude of these effects estimated by Cohen's $d$ ranged from medium to large: OO/MO = 0.61, OO/YO = 1.06, OO/Y = 1.57, MO/YO = 0.52, MO/Y = 1.15, and YO/Y = 0.70. In the case of CRT, the OO group was more variable than all other groups. Both the MO and YO groups were more inconsistent than the Y group, but the MO and YO did not differ significantly. The magnitude of the significant effects again ranged from medium to large: OO/MO = 0.54, OO/YO = 0.78, OO/Y = 1.69, MO/Y = 1.19, and YO/Y = 0.95. For
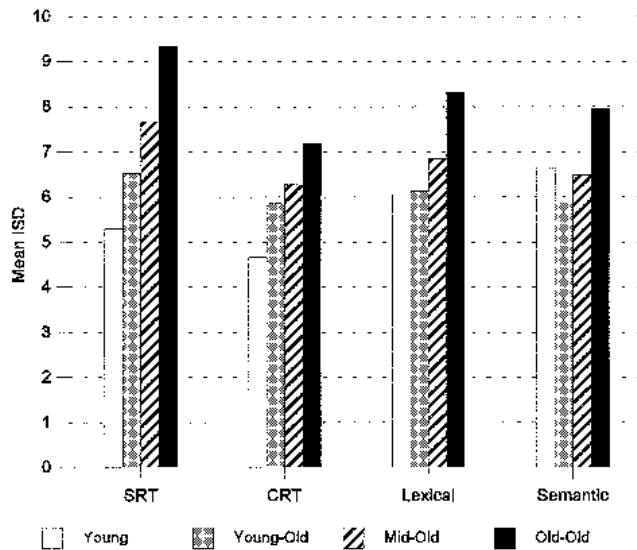
Figure 3. Mean latency intraindividual standard deviation (ISD) scores by age group for four reaction time tasks. SRT = simple reaction time; CRT = choice reaction time.

the lexical task, the OO group showed significantly more intraindividual variability than all other groups. The MO group was more variable than the YO and Y groups. However, the Y and YO groups did not differ significantly. Cohen's $d$ ranged from small to large: OO/MO = 0.55, OO/YO = 0.89, OO/Y = 0.95, MO/YO = 0.31, and MO/Y = 0.38. Finally, in the case of the semantic task, the OO group was again more inconsistent than all other age groups. The MO group exhibited more intraindividual variability in performance than the YO group, but it did not differ significantly from the Y group. Interestingly, the Y group was significantly more inconsistent than the YO group. For the semantic task, the magnitude of the differences ranged from small to large: OO/MO = 0.75, OO/YO = 1.10, OO/Y = 0.65, MO/YO = 0.34, and YO/Y = 0.39.

The significant gender effect was the result of a difference on the SRT task alone, $F(1, 854) = 12.37, p < .001, \eta^2 = .01$. Women ($M = 7.96$) showed slightly more intraindividual variability on this task than men ($M = 7.19$). The magnitude of this effect ($d$) was small (0.27).

We performed the same analyses using the coefficient of variation, which provides a measure of intraindividual variation relative to the individual's own mean score. The substantive results were identical to those found with the ISD measure. Finally, we repeated all of the analyses covarying education and observed the same pattern of significant results.

*Percentile analysis.*—An alternative means of examining inconsistency (Salthouse, 1993) focuses on the question of whether RT slopes, plotted as a function of age, are symmetric across the entire distribution of RT latencies (e.g., 10th, 25th, 50th, 75th, and 90th percentiles). Salthouse argued that two key patterns should be observed if slow RTs are influenced by the same processes that affect fast RTs: (a) age-related slopes for different percentiles of the distri-

bution should be symmetric and (b) age should share considerable variance with both low and high percentiles of the distribution. To examine intraindividual variability as a function of RT distributions, we computed ISDs (using purified residual scores) separately for the 20th and 80th percentiles for each of the four RT tasks. Specifically, each individual's distribution of RT scores was sorted in ascending order, 20th and 80th percentile cutoffs were determined, and two ISD estimates were calculated for each individual for each task (one estimate reflecting variability for fast trials below the 21st percentile and one estimate for slow trials above the 79th percentile). Higher ISD scores reflect more inconsistent individual performances within each percentile range.

A 4 (age) $\times$ 2 (gender) $\times$ 2 (percentile) repeated measures ANOVA was computed using the 20th and 80th-percentile ISD scores for each of the four tasks. Multivariate tests revealed significant omnibus effects associated with age, Wilks's $\lambda = .707, F(12, 2559) = 26.27, p < .001, \eta^2 = .11$; gender, Wilks's $\lambda = .988, F(4, 851) = 2.63, p < .05, \eta^2 = .01$; percentile, Wilks's $\lambda = .176, F(4, 851) = 999.33, p < .001, \eta^2 = .82$; Age $\times$ Percentile, Wilks's $\lambda = .805, F(12, 2559) = 16.03, p < .001, \eta^2 = .07$; and Gender $\times$ Percentile, Wilks's $\lambda = .988, F(4, 851) = 2.55, p < .05, \eta^2 = .01$. No additional interactions were significant.

Qualifying individual main effects of Age and Percentile, univariate ANOVAs indicated there were significant Age $\times$ Percentile interactions for each of the four tasks: SRT: $F(3, 854) = 13.43, p < .001, \eta^2 = .05$; CRT: $F(3, 854) = 11.52, p < .001, \eta^2 = .04$; lexical: $F(3, 854) = 24.25, p < .001, \eta^2 = .08$; and semantic: $F(3, 854) = 18.54, p < .001, \eta^2 = .06$. Figure 4 shows mean ISDs as a function of age group for the 20th- and 80th-percentile trials. Age differences in inconsistency were substantially more pronounced for the slowest compared with the fastest segments of the RT distribution. A univariate gender effect for the SRT task alone was qualified by a significant Gender $\times$ Percentile interaction, $F(1, 854) = 4.75, p < .05, \eta^2 = .01$. For 20th percentile RTs, variability estimates for women ($M = 1.49$) and men ($M = 1.44$) were equivalent. For 80th-percentile RTs, women ($M = 7.37$) showed slightly more inconsistency on this task than men ($M = 6.52$). No other significant effects associated with gender were found.

Finally, following Salthouse (1993), we used a regression approach to examine whether age shared considerable variance with variability estimates for both low and high percentiles of the distribution. For each task, hierarchical regressions were computed using age and 20th percentile ISDs as predictors of 80th-percentile ISDs. Age was a significant predictor ($p < .01$) of 80th-percentile ISDs for all tasks (SRT: $R^2 = .049$; CRT: $R^2 = .063$; lexical: $R^2 = .054$; and semantic: $R^2 = .021$). However, partialing out variability in 20th-percentile RTs had little effect on the relation of age to variability in 80th-percentile RTs. Less than 2% of the age-related variance was attenuated for CRT and lexical tasks. Variability in 20th-percentile RTs actually served a suppressor function for both the SRT and semantic tasks. The association between age and 80th-percentile ISDs increased marginally after partialing out variability in 20th-percentile ISDs.
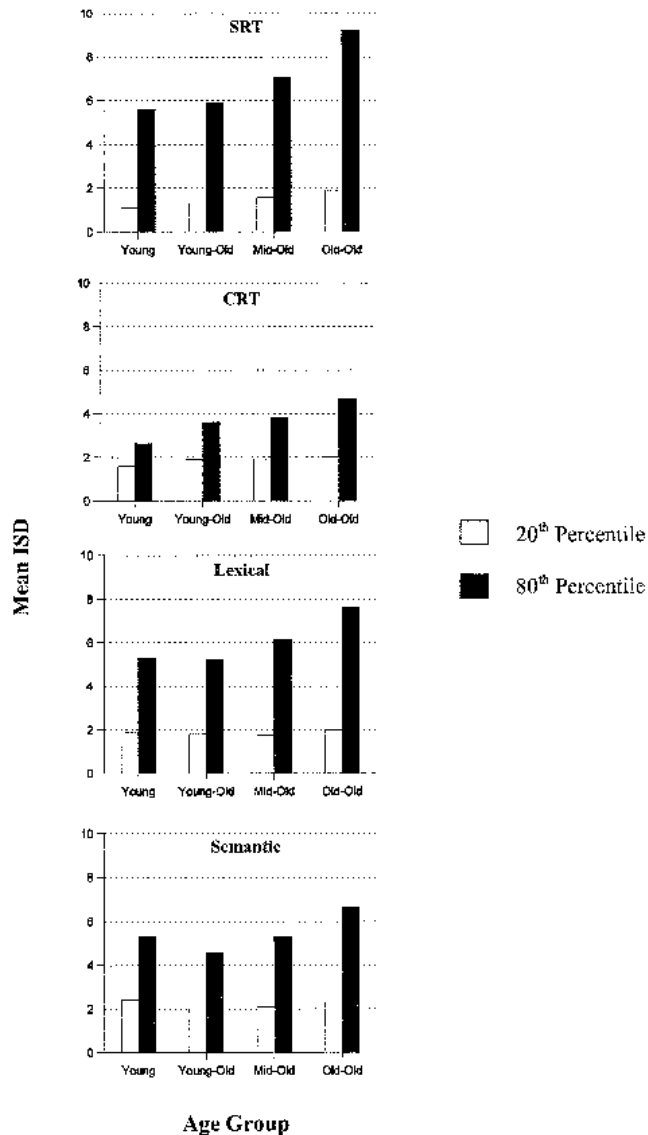
Figure 4. Mean latency intraindividual standard deviation (ISD) scores for faster (20th percentile) and slower (80th percentile) responses by age group for four reaction time tasks. SRT = simple reaction time; CRT = choice reaction time.

*Quantile-quantile plots.*—The previous analysis suggests that age differences in inconsistency are greater for slower RTs and imply that the RT distributions are more positively skewed with increasing age. Standard quantile-quantile plots (Q-Q plots), a graphical technique for determining if two data sets come from a common distribution, represent one means of examining multiple parameter estimates of RT distributions from two separate groups (Ratcliff, Spieler, & McKoon, 2000). Recent findings have suggested that Q-Q plots represent a rich source of evidence about age differences in variability (Maylor & Rabbitt, 1994; Ratcliff et al., 2000). Specifically, Ratcliff and colleagues demonstrated that the slope of the Q-Q plot represents a ratio of the standard deviation for one group relative to the standard deviation of the other. A linear plot indicates the under-

lying distributions are from the same family (e.g., normal), with deviations from linearity pinpointing the area of the distributions where the two groups differ most. Q-Q plots are most commonly constructed using raw RTs. In the present analysis, we used Q-Q plots to compare distributions of ISD scores. We did this because ISDs, as we have calculated them, are statistically independent of group mean-level performance. This analysis, then, allows us to examine the robustness and distribution characteristics of intraindividual variability after controlling for mean/standard deviation confounds. Figure 5 shows plots of ISD scores for the OO group as a function of ISD scores for the Y group (see Appendix, Note 3). For both SRT and CRT, the slopes indicate that (a) the distributions of ISD scores (inconsistency) for the older adults were wider than those of the younger adults (slope greater than 1.0) and (b) the distributions of the OO group were more positively skewed, albeit subtly, for CRT (see histograms on the vertical axes). The plot for the lexical task shows that the OO group were 1.4 times more inconsistent than the Y group, with the straight line indicating that both groups had similar ISD distributions. The plot for the semantic task suggests that the OO group was only slightly more inconsistent than the Y group. Moreover, the departure of the upper quantiles of the Q-Q plot from a linear function suggests that the distribution of the Y group was actually more positively skewed than that of the OO group (see histogram on the horizontal axis).

## INTERCORRELATIONS

Table 3 shows the intercorrelations among the various measures of within-person variability—the single indicator of intraindividual variability across tasks (dispersion) and the indicators of intraindividual variability across trials for each of the four tasks (inconsistency). The correlations are low in magnitude, although they are statistically significant given the large sample size. Interestingly, however, all of the relationships are positive. This indicates that (a) individuals who were more variable across tasks (had a more dispersed profile) were also more variable across time (showed inconsistent performance across trials) and (b) individuals who were more variable across trials on one RT task were also more variable across trials on the other RT tasks. We also examined these relationships separately by age group. Although the number of significant values varied because of group differences in sample size, examination of Fisher's *z* indicated few significant differences in correlations across the age groups.

## RELATIONSHIPS TO COGNITIVE PERFORMANCE

We began our examination of whether individual differences in inconsistency of RT latency were predictive of mean performance (accuracy) on other cognitive tasks by computing zero-order correlations among the two sets of measures. As seen in Table 4, greater inconsistency in RT performance was associated with poorer performance on the cognitive composites. The table shows that these relationships were somewhat more widespread in the oldest compared with the youngest group. For example, significant correlations for the youngest group were observed largely between intraindividual variability in the two verbal RT
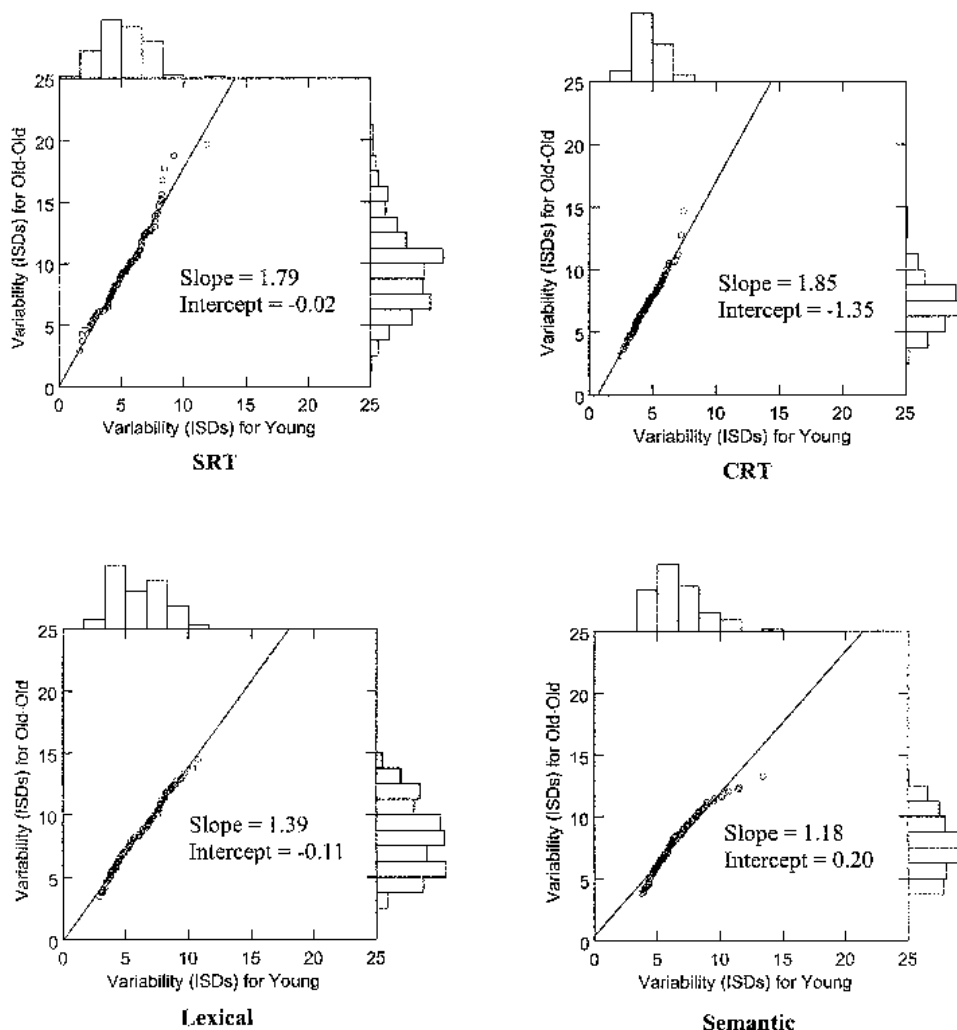
*HULTSCH ET AL.*



Figure 5. Quantile-quantile plots for intraindividual variability scores for the Old-Old group as a function of the intraindividual variability scores for the Young group. ISD = intraindividual standard deviation; SRT = simple reaction time; CRT = choice reaction time.

tasks and perceptual speed and episodic memory. In contrast, with increasing age the number of significant relationships tended to increase. In the oldest group, significant correlations were observed between all measures of intraindividual variability and cognitive performance, with one exception.

However, a similar pattern of zero-order correlations was observed between overall speed of performance (mean latency) on the four RT tasks and performance on the same cognitive composites. Therefore, it is important to examine unique and shared contributions of ISD estimates and mean estimates as predictors of cognitive performance. To the extent that intraindividual variability represents an influential and independent marker of cognitive function, ISD estimates should account for a significant proportion of variance in cognitive performance over and above mean-level influences.

We used partial set correlation (Cohen, 1982) to examine the unique and shared influences of unpurified intraindividual mean (IM) and purified ISD estimates as predictors of cognition. Partial set correlation permits examination of as-

sociations among criterion and predictor constructs that are identified by multiple measures. Variance for each of the four cognitive composites was partitioned into that uniquely associated with IMs and ISDs as well as variance shared between them. For each analysis, the dependent variables consisted of mean performance on the cognitive tasks making up the specific cognitive composite (e.g., episodic memory: word recall, story recall). The independent variables con-

Table 3. Intercorrelations of Measures of Within-Person Variability in Reaction Time

| Measure | ISD (tasks) | ISD (trials) | | | |
|---|---|---|---|---|---|
| | | SRT | CRT | Lexical | Semantic |
| Simple reaction time (SRT) | .21*** | — | | | |
| Choice reaction time (CRT) | .12** | .29*** | — | | |
| Lexical | .26*** | .31*** | .20*** | — | |
| Semantic | .22*** | .24*** | .17*** | .47*** | — |

*Note*: ISD = intraindividual standard deviation.

**p < .01; ***p < .001.

Table 4. Correlations of Intraindividual Standard Deviation (ISD) Across Trials on Four Reaction Time (RT) Tasks With Mean Performance on Other Cognitive Measures by Age Group

| Age Group | RT–ISD (trials) | | | |
|---|---|---|---|---|
| Cognitive Measure | SRT | CRT | Lexical | Semantic |
| **Young** | | | | |
| Perceptual speed | −.15 | −.20* | −.41** | −.33** |
| Working memory | −.02 | −.09 | −.16 | −.11 |
| Episodic memory | −.14 | −.15 | −.31** | −.48** |
| Crystallized ability | −.11 | −.04 | −.13 | −.14 |
| **Young-Old** | | | | |
| Perceptual speed | −.23** | −.05 | −.17* | −.15* |
| Working memory | −.20** | −.12 | −.15* | −.19* |
| Episodic memory | −.14 | .00 | −.15* | −.18* |
| Crystallized ability | −.11 | .02 | −.22** | −.28** |
| **Mid-Old** | | | | |
| Perceptual speed | −.17** | −.17** | −.26** | −.20** |
| Working memory | −.10 | −.07 | −.21** | −.10 |
| Episodic memory | −.04 | −.08 | −.20** | −.18** |
| Crystallized ability | −.06 | .03 | −.17** | −.17** |
| **Old-Old** | | | | |
| Perceptual speed | −.35** | −.34** | −.42** | −.27** |
| Working memory | −.23** | −.15* | −.31** | −.28** |
| Episodic memory | −.21** | −.21** | −.29** | −.24** |
| Crystallized ability | −.24** | −.13 | −.28** | −.28** |

*Note*: SRT = simple reaction time; CRT = choice reaction time.
*$p < .05$; **$p < .01$.

Table 5. Relative Contribution of Intraindividual Means (IMs) and Intraindividual Standard Deviations (ISDs) as Predictors of Cognitive Performance

| Variable | Multivariate $R^2$ (Total) | Multivariate Partial $R^2$ | | |
|---|---|---|---|---|
| | | Shared | Unique IM | Unique ISD |
| **Perceptual speed** | | | | |
| Nonverbal RT | .533** | .124 (23.3) | .321 (60.2)** | .088 (16.5)** |
| Verbal RT | .331** | .152 (45.9) | .171 (51.7)** | .008 (2.4) |
| **Working memory** | | | | |
| Nonverbal RT | .224** | .098 (43.8) | .102 (45.5)** | .024 (10.7)** |
| Verbal RT | .174** | .110 (63.2) | .056 (32.2)** | .008 (4.6) |
| **Episodic memory** | | | | |
| Nonverbal RT | .201** | .082 (40.8) | .095 (47.3)** | .024 (11.9)** |
| Verbal RT | .264** | .122 (46.2) | .127 (48.1)** | .015 (5.7) |
| **Crystallized ability** | | | | |
| Nonverbal RT | .133** | .061 (45.9) | .045 (33.8)** | .027 (20.3)** |
| Verbal RT | .188** | .058 (30.8) | .118 (62.8)** | .012 (6.4) |

*Notes*: Values in parentheses represent percentage of total multivariate $R^2$ accounted for. Estimates of shared variance were derived by subtracting unique IM and ISD estimates from total multivariate $R^2$. RT = reaction time.
**$p < .01$.

sisted of IMs and ISDs of two classes of RT tasks: nonverbal RT (SRT and CRT) and verbal RT (lexical and semantic) to examine whether patterns of prediction varied as a function of type of RT measure. Three set correlations were computed: regression of cognitive measures onto IM and ISD without partialing any variables, regression of cognitive measures onto ISD partialing out IM performance, and regression of cognitive measures onto IM performance partialing out ISD.

Table 5 shows the amount of total variance (i.e., variance in cognitive performance predicted by both IMs and ISDs) that is uniquely accounted for by nonverbal and verbal IMs and ISDs, as well as the variance shared between these predictors. Performance for each cognitive domain was significantly predicted by both IMs and ISDs (total) for nonverbal and verbal RT. Not surprisingly, mean and variability estimates shared a considerable amount of overlapping variance (shared) as predictors of total $R^2$. However, as expected, both mean and variability estimates also demonstrated unique predictive contributions. For both categories of RT measures, mean-level performance (unique IM) significantly predicted performance in cognition independent of ISDs. Mean-level performance uniquely accounted for between 34% and 60% of total $R^2$ for nonverbal RT and 32% to 63% of total $R^2$ for verbal RT. Of particular interest, intraindividual variability estimates (unique ISD) for nonverbal RT significantly predicted variance in cognitive performance over and above mean-level influences for each of the cognitive domains examined. ISDs for nonverbal RT uniquely accounted for between 11% (working memory) to 20% (crystallized ability) of total $R^2$. In contrast, ISDs in verbal RT did not account for any unique variance in cogni-

tive performance over and above mean-level influences. These results indicate that patterns of prediction differ as a function of type of RT measure and not as a function of cognitive domain assessed.

## Discussion

This article represents the first effort to examine age differences in all three basic types of variability within the same data set. We measured variability between persons (diversity) and two types of variability within persons (dispersion of performance across tasks and inconsistency of performance across trials) on four RT tasks. As hypothesized, we found that all three types of variability were greater in older as compared with younger adults. Importantly, we observed these significant age differences in variability even after statistical control of group differences in speed of performance. Thus, the differences cannot be attributed to an artifact of the relationship between the mean and the standard deviation.

Consistent with previous reports (e.g., Christensen et al., 1994; Morse, 1993), we found our older participants showed greater diversity in RT performance than younger adults. When the youngest age group was included in the analysis, the results showed evidence for a significant quadratic as well as a linear trend, suggesting an increasing degree of diversity in the oldest participants. However, analyses conducted with the three older adult age groups showed only significant linear effects for the four tasks. This suggests the quadratic effect was largely a function of the extreme-groups comparison. Nevertheless, we observed increasing diversity in performance from the mid-50s through the late 80s. Two other points are worth noting with respect to age differences in diversity. First, although interindividual differences increased across the age groups, the magnitude of the differences was relatively small. In most cases, the linear age effect accounted for around 2% of the variance. Second, the tasks used in the present study were

driven largely by relatively basic processing mechanisms. Other studies have shown that age-related differences in diversity are less likely to be observed with measures that focus on the assessment of acquired knowledge (e.g., Christensen et al., 1994).

We also observed evidence for increasing dispersion of scores with increasing age. That is, on average, older adults had more uneven intraindividual profiles of performance across the four RT tasks than younger adults. Significant differences were observed between the Y group and the two oldest groups and also between the YO group and the two older groups. Interestingly, these age differences in dispersion were found despite the relative similarity of the four tasks. However, the magnitude of the differences as indicated by Cohen's *d* was relatively modest, ranging from small to medium according to his convention. Our results are in agreement with Christensen, Mackinnon, Korten, Jorm, Henderson, and Jacomb (1999), who also observed increasing dispersion of scores across measures of speed, memory, and spatial functioning. In contrast, Lindenberger and Baltes (1997) reported no age differences in dispersion across 14 measures of intelligence for higher ability adults and actually observed decreases in dispersion for lower ability adults. In part, these discrepancies may be a result of the older age of the sample used by Lindenberger and Baltes (70–103 years). Increases in dispersion may be seen in younger and relatively healthy older adults followed by decreases in dispersion very late in life as more individuals experience significant cognitive decline in all domains associated with the end of life.

We found evidence for substantial age differences in inconsistency of RT performance across trials. In general, older adults showed greater intraindividual variability in RT latency across trials than younger adults on all four tasks. Age differences were particularly pronounced for individuals age 75 and above. Participants in this group showed greater inconsistency than all other age groups on all tasks. The effect sizes associated with the age group differences in intraindividual variability were sizable, ranging from medium to large in most cases. This pattern of results is consistent with a number of other recent analyses (Rabbitt, 2000; Spieler et al., 1996; West & Baylis, 1998; West et al., in press).

Our results also suggest that age differences in inconsistency vary across the RT distribution. Age differences in intraindividual variability were larger for slower (80th-percentile) than faster (20th-percentile) responses, although age differences were significant in both cases. Moreover, in contrast to Salthouse (1993), we found that partialing out age-related influences from the fastest responses had little effect on the relation of age to variability for the slowest responses. This result is consistent with the argument that older adults may experience temporary lapses of attention (Bunce, Warr, & Cochrane, 1993) or executive control (West et al., in press) that contribute to greater inconsistency of performance. Our analysis of age group differences for the fastest and slowest responses implied, but did not demonstrate, that the response distributions of intraindividual variability were more positively skewed with increasing age. The Q-Q plots demonstrated age group differences in

intraindividual variability as well as the shape of distributions for some tasks. The differences in slope confirmed the findings of our earlier analysis of purified residuals; all tasks showed slopes greater than 1.0 indicating the older group was more inconsistent than the younger group. Importantly, the plots of mean-independent ISDs showed age group and task differences in the tails of the distributions. In the case of SRT and to a lesser extent CRT, the distributions for the older adults were positively skewed, whereas the distributions for the younger adults were relatively normal. In contrast, there was little evidence for group differences in the distribution for the lexical task, and in the case of the semantic task, it was the distribution of the younger group that showed greater positive skew.

Ordinarily, one might expect to observe larger age differences on more complex tasks compared with more simple tasks (e.g., Salthouse, 1991; West et al., in press). One plausible interpretation of the differences we observed between the nonverbal and verbal RT tasks is related to the greater verbal facility of older adults compared with younger adults (e.g., Wingfield & Stine-Morrow, 2000). In the case of the verbal tasks, the older adults' verbal ability may have provided them with access to compensatory mechanisms that reduced their relative inconsistency to some degree. Consistent with this view, Hale and Myerson (1996) have explicitly shown age-related slowing to be much larger in the nonlexical domain than in the lexical domain. Nevertheless, compensatory mechanisms may be more easily invoked for multicomponent tasks that rely on knowledge than for more primitive tasks that rely on speed (Dixon & Bäckman, 1999). In domains where compensatory mechanisms are less easily activated, age differences in inconsistency may increase with task difficulty. For example, West and colleagues (in press) found exactly this pattern for RT tasks that varied in their executive demand.

The present data cannot be used to identify the specific mechanisms that may underlie increased variability with increasing age. Li and Lindenberger (1999) have suggested that both age-related decreases in level of performance and increases in interindividual differences in performance may be driven by increasing intraindividual variability in neurobiological mechanisms in the brain. This view suggests that measures of intraindividual variability may be a plausible behavioral indicator of aging-induced deterioration of general neurobiological mechanisms that compromise the integrity of the brain across a wide range of areas and functional circuitry. In particular, it has been suggested that older brains may need to recruit additional resources to manage executive functions of otherwise relatively simple tasks (Dixon & Bäckman, 1999). Thus, even localized neural deficits may be expressed as a generalized impairment (Raz, 2000). Such changes have been hypothesized as the common cause for aging-associated losses in cognitive capacity and plasticity (Baltes & Lindenberger, 1997; Lindenberger & Baltes, 1994).

In our data, the potential involvement of neurological mechanisms in producing age differences in intraindividual variability is suggested by the relative consistency of individual differences in measures of within-person variability across tasks. The correlational analyses indicated consistent

patterns of correlations among (a) the measure of dispersion and measures of inconsistency and (b) among measures of inconsistency across tasks. That is, there was a consistent positive manifold to intercorrelations of the various indicators of within-person variability. Individuals who showed greater dispersion across the four tasks also tended to show greater inconsistency across trials on all four RT tasks. Similarly, individuals who showed greater inconsistency across trials on one task tended to show greater inconsistency on the other tasks as well. Other recent studies have shown positive correlations between intraindividual variability measured across trials and across occasions (Hultsch et al., 2000; Rabbitt et al., 2001). The magnitude of within-person variability, then, appears to be somewhat characteristic of the individual—both across tasks and over time. This is what one would expect to find if such variability were substantially influenced by relatively stable endogenous mechanisms such as neurological dysfunction rather than relatively labile exogenous influences such as pain, fatigue, and stress.

Regardless of the specific mechanisms involved, the present data indicate that intraindividual variability in RT performance is negatively correlated with level of performance on a wide range of more complex cognitive tasks. In general, these relationships tend to be more widespread with increasing age. The partial set correlation analysis also indicated that intraindividual variability in nonverbal RT latency (i.e., variability for SRT and CRT tasks) was a unique predictor of cognitive performance. These nonverbal RT measures of intraindividual variability uniquely accounted for between 11% and 20% of the variance in cognitive performance independent of mean-level influences. The unique contribution of intraindividual variability as a predictor of performance across a continuum of cognitive tasks is what one would expect if inconsistency is influenced by relatively stable endogenous mechanisms. Interestingly, intraindividual variability in verbal RT performance did not uniquely account for variance in cognitive performance (range from 2% to 6%) over and above verbal RT means. This finding is also consistent with the verbal facility hypothesis suggesting that, for verbal RT tasks, older adults exhibit less inconsistency relative to their performance on nonverbal RT tasks as their verbal facility serves a compensatory function.

In summary, the present results indicate that older adults are more variable than younger adults on all three types of variability measured. In particular, the results point to intraindividual variability or inconsistency in response speed as a potentially important predictor of cognitive performance independent of overall age differences in speed. Other recent studies have been in agreement with this general view (e.g., Hultsch et al., 2000; Rabbitt et al., 2001; West et al., in press). How should this consistent pattern of age differences be interpreted? Nesselroade (1991) noted that intraindividual variability contributes to an individual's performance at any one point in time, thus confounding measures of inconsistency and diversity. However, Nesselroade and Li and Lindenberger (1999) have also suggested that characteristics of variability (e.g., inconsistency) at the level of individuals can be a source of both stable interindividual differences and intraindividual change in performance. Thus, from this perspective, the linkage between within- and between-person variability may be seen as substantively important rather than as a methodological confound. In effect, intraindividual variability may influence both diversity and dispersion as well as long-term developmental change.

The available results are largely descriptive and need to be expanded. Further analysis of age and task differences in the shape of response time distributions can provide important insights into the operation of cognitive processes (Ratcliff et al., 2000). Our finding that there appeared to be differences in the shapes of the ISD distributions suggests that future research should consider distribution parameters for this indicator across tasks. We also suggest that a particularly important question is whether intraindividual variability in performance is predictive of cognitive change over time rather than simply level of performance as already suggested by extant data. Examining this issue with longitudinal as well as cross-sectional data is particularly important because recent research has shown important dissociations between cross-sectional results both at the level of means and relationships among variables (Hultsch et al., 1998). If intraindividual variability is an independent predictor of actual changes in cognitive functioning, it will strengthen the argument that inconsistency represents a potentially useful indicator of cognitive aging from both a theoretical and a practical perspective.

REFERENCES

Alwin, D. F. (1994). Aging, personality, and social change: The stability of individual differences over the adult life span. In D. L. Featherman, R. M. Lerner, & M. Perlmutter (Eds.), *Life-span development and behavior* (Vol. 12, pp. 135–185). Hillsdale, NJ: Erlbaum.

Anstey, K. J. (1999). Sensorimotor and forced expiratory volume as correlates of speed, accuracy, and variability in reaction time performance in late adulthood. *Aging, Neuropsychology, and Cognition, 6*, 84–95.

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult lifespan: A new window to the study of cognitive aging? *Psychology and Aging, 12*, 12–21.

Bunce, D. J., Warr, P. B., & Cochrane, T. (1993). Blocks in choice responding as a function of age and physical fitness. *Psychology and Aging, 8*, 26–33.

Cattell, R. B. (1966). The data box: Its ordering of total resources in terms of possible relational systems. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 67–128). Chicago: Rand McNally.

Christensen, H., Mackinnon, A. J., Jorm, A. F., Henderson, A. S., Scott, L. R., & Korten, A. E. (1994). Age differences and interindividual vari-

ation in cognition in community-dwelling elderly. *Psychology and Aging, 9,* 381–390.

Christensen, H., Mackinnon, A. J., Korten, A. E., Jorm, A. F., Henderson, A. S., & Jacomb, P. (1999). Dispersion in cognitive ability as a function of age: A longitudinal study of an elderly community sample. *Aging, Neuropsychology, and Cognition, 6,* 214–228.

Christensen, H., Mackinnon, A. J., Korten, A. E., Jorm, A. F., Henderson, A. S., Jacomb, P., & Rodgers, B. (1999). An analysis of diversity in the cognitive performance of elderly community dwellers: Individual differences in change scores as a function of age. *Psychology and Aging, 14,* 365–379.

Cohen, J. (1982). Set correlation as a general multivariate data-analytic method. *Multivariate Behavioral Research, 17,* 301–341.

Dixon, R. A., & Bäckman, L. (1999). Principles of compensation in cognitive neurorehabilitation. In D. T. Stuss, G. Winocur, & I. H. Robertson (Eds.), *Cognitive neurorehabilitation* (pp. 59–72). Cambridge, England: Cambridge University Press.

Dixon, R. A., Hertzog, C., Friesen, I. C., & Hultsch, D. F. (1993). Assessment of intraindividual change in text recall of elderly adults. In H. H. Brownell & Y. Joanette (Eds.), *Narrative discourse in neurologically impaired and normal aging adults* (pp. 77–101). San Diego, CA: Singular.

Dixon, R. A., Hultsch, D. F., & Hertzog, C. (1989). *A manual of three-tiered structurally equivalent texts for use in aging research* (CRGCA Tech. Rep. No. 2). Victoria, British Columbia, Canada: University of Victoria, Department of Psychology.

Dixon, R. A., Wahlin, Å., Maitland, S. B., Hultsch, D. F., Hertzog, C., & Bäckman, L. (in press). Episodic memory change in late adulthood: Generalizability across samples and performance indices. *Memory & Cognition.*

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Fozard, J. L., Vercruyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore Longitudinal Study of Aging. *Journal of Gerontology: Psychological Sciences, 49,* P179–P189.

Gergen, K. J. (1977). Stability, change, and chance in understanding human development. In N. Datan & H. W. Reese (Eds.), *Life-span developmental psychology: Dialectical perspectives on experimental research* (pp. 135–158). New York: Academic Press.

Gordon, B., & Carson, K. (1990). The basis for choice reaction time slowing in Alzheimer's disease. *Brain and Cognition, 13,* 148–166.

Hale, S., & Myerson, J. (1996). Experimental evidence for differential slowing in the lexical and nonlexical domains. *Aging, Neuropsychology, and Cognition, 3,* 154–165.

Hale, S., Myerson, J., Smith, G. A., & Poon, L. W. (1988). Age, variability and speed: Between-subjects diversity. *Psychology and Aging, 3,* 407–410.

Hendrickson, A. E. (1982). The biological basis of intelligence. Part I: Theory. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 151–196). Berlin: Springer-Verlag.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1992). Intraindividual change in text recall of the elderly. *Brain and Language, 42,* 248–269.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 598–615.

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

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory change in the aged.* New York: Cambridge University Press.

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14,* 588–598.

Jensen, A. R. (1982). Reaction time and psychometric *g.* In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). Berlin: Springer-Verlag.

Li, S.-C., Aggen, S. H., Nesselroade, J. R., & Baltes, P. B. (2001). Short-term fluctuations in elderly people's sensorimotor functioning predicts text and spatial memory performance: The MacArthur successful aging studies. *Gerontology, 47,* 100–116.

Li, S.-C., & Lindenberger, U. (1999). Cross-level unification: A computa-

tional exploration of the link between deterioration of neurotransmitter systems and dedifferentiation of cognitive abilities in old age. In L.-G. Nilsson & H. Markowitsch (Eds.), *Cognitive neuroscience and memory* (pp. 103–146). Toronto: Hogrefe & Huber.

Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging, 9,* 339–355.

Lindenberger, U., & Baltes, P. B. (1997). Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study. *Psychology and Aging, 12,* 410–432.

Maylor, E. A., & Rabbitt, P. M. A. (1994). Applying Brinley plots to individuals: Effects of aging on performance distributions in two speeded tasks. *Psychology and Aging, 9,* 224–230.

Morse, C. K. (1993). Does variability increase with age? An archival study of cognitive measures. *Psychology and Aging, 8,* 156–164.

Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review, 97,* 475–487.

Nelson, A. E., & Dannefer, D. (1992). Aged heterogeneity: Fact or fiction? The fate of diversity in gerontological research. *The Gerontologist, 32,* 17–23.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19,* 338–368.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 92–105). Washington, DC: American Psychological Association.

Nesselroade, J. R., & Boker, S. M. (1994). Assessing constancy and change. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 121–147). Washington, DC: American Psychological Association.

Nesselroade, J. R., & Featherman, D. L. (1997). Establishing a reference frame against which to chart age-related changes. In M. A. Hardy (Ed.), *Studying aging and social change: Conceptual and methodological issues* (pp. 191–205). Newbury Park, CA: Sage.

Nesselroade, J. R., & Ford, D. H. (1985). P-technique comes of age: Multivariate, replicated, single-subject designs for research on older adults. *Research on Aging, 7,* 46–80.

Rabbitt, P. (1993). Does it all go together when it goes? The 19th Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 46A,* 385–434.

Rabbitt, P. M. A. (2000). Measurement indices, functional characteristics, and psychometric constructs in cognitive aging. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 160–187). New York: Oxford University Press.

Rabbitt, P., Osman, P., & Moore, B. (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *The Quarterly Journal of Experimental Psychology, A, 54,* 981–1003.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.

Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin and Review, 7,* 1–25.

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 1–90). Mahwah, NJ: Erlbaum.

Rowe, J. W., & Kahn, R. L. (1997). Successful aging. *The Gerontologist, 37,* 433–440.

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

Salthouse, T. A. (1993). Attentional blocks are not responsible for age-related slowing. *Journal of Gerontology: Psychological Sciences, 48,* P263–P270.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763–776.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study.* New York: Cambridge University Press.

Shammi, P., Bosman, E., & Stuss, D. T. (1998). Aging and variability in performance. *Aging, Neuropsychology, and Cognition, 5,* 1–13.

Slifkin, A. B., & Newell, K. M. (1998). Is variability in human perfor-

mance a reflection of system noise? *Current Directions in Psychological Science*, *7*, 170–177.

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 461–479.

Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of resonse time distributions. *Journal of Experimental Psychology: Human Performance and Perception*, *26*, 506–526.

Stuss, D. T., Pogue, J., Buckle, L., & Bondar, J. (1994). Characterization of stability of performance in patients with traumatic brain injury: Variability and consistency on reaction time tests. *Neuropsychology*, *8*, 316–324.

Welford, A. T. (1980). Relationships between reaction time and fatigue, stress, age and sex. In A. T. Welford (Ed.), *Reaction times* (pp. 321–354). New York: Academic Press.

West, R., & Baylis, G. C. (1998). Effects of increased response dominance and contextual disintegration on the Stroop interference effect in older adults. *Psychology and Aging*, *13*, 206–217.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (in press). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition.*

Wingfield, A., & Stine-Morrow, E. A. L. (2000). Language and speech. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 359–416). Mahwah, NJ: Erlbaum.

## Appendix

### *Notes*

1. This definition assumes each task consists of a single condition. This type of variability could also be defined as variability associated with multiple conditions of a single task.

2. Cohen's *d* is calculated as the difference between the means (in original metric) divided by the average standard deviation of both groups. The result is an index devoid of arbitrary scaling metric that expresses the distance between means in units of variability.

3. Given the linear age patterns observed in Figure 4 and space limitations, we compared only the two extreme age groups. An analysis plotting all older adults as a function of the young adults yielded a similar but attenuated pattern of results.