

# On the Relation of Mean Reaction Time and Intraindividual Reaction Time Variability

Florian Schmiedek

Max Planck Institute for Human Development  
and Humboldt-Universität

Martin Lövdén

Max Planck Institute for Human Development  
and Lund University

Ulman Lindenberger

Max Planck Institute for Human Development

Researchers often statistically control for means when examining individual or age-associated differences in variances, assuming that the relation between the 2 is linear and invariant within and across individuals and age groups. We tested this assumption in the domain of working memory by applying variance-heterogeneity multilevel models to reaction times in the  $n$ -back task. Data are from the COGITO study, which comprises 101 younger and 103 older adults assessed in over 100 daily sessions. We found that relations between means and variances vary reliably across age groups and individuals, thereby contradicting the invariant linearity assumption. We argue that statistical control approaches need to be replaced by theoretical models that simultaneously estimate central tendency and dispersion of latencies and accuracies and illustrate this claim by applying the diffusion model to the same data. Finally, we note that differences in reliability between estimates for means and variances need to be considered when comparing their unique contributions to developmental outcomes.

*Keywords:* working memory, intraindividual variability, multilevel models, variance heterogeneity models, reliability

The importance of investigating intraindividual variability has been acknowledged in psychology for over half a century. From the very beginning, researchers hypothesized that intraindividual fluctuations in behavior may (a) be separable from measurement error, (b) differ in magnitude between individuals, and (c) predict other psychological phenomena (Fiske & Rice, 1955). Taking a

multivariate perspective, Cattell (1952) called for a broadening of theoretical concepts, research designs, and analysis methods to include structured relations among constructs within persons (Nesselroade, 1984). Combining these ideas with lifespan developmental conceptions, Nesselroade (1991) underscored the importance of incorporating concepts of short-term variability in theories and studies of long-term change. Further calls for including idiographic perspectives in investigations of psychological and developmental mechanisms have been made (Borsboom, Mellenbergh, & van Heerden, 2003; Molenaar, 2004; Molenaar & Campbell, 2009).

Following these calls, researchers have begun to conduct empirical studies that examine the commonalities and the differences between the structures representing within-person variations and the structures representing between-person differences (for reviews, see Hultsch & MacDonald, 2004; Lindenberger & von Oertzen, 2006). In the domain of cognitive functioning, a few empirical studies have invested the effort of testing many individuals over many trials and occasions to compare interindividual and intraindividual variability. Intraindividual variability in cognitive performance can be investigated for different domains of cognitive functioning (e.g., perceptual speed, episodic memory, or working memory), at different time scales (e.g., responses on single trials or performance means on daily testing occasions), and using different measures of performance (e.g., accuracy and reaction time). Because the empirical part of this article is on the relation between intraindividual means (iM) and intraindividual standard deviations (iSD) of reaction times (RT), the following literature review focuses on intraindividual RT variability.

---

Florian Schmiedek, Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany; and Institute of Psychology, Humboldt-Universität, Berlin, Germany. Martin Lövdén, Center for Lifespan Psychology, Max Planck Institute for Human Development; and Department of Psychology, Lund University, Lund, Sweden. Ulman Lindenberger, Center for Lifespan Psychology, Max Planck Institute for Human Development.

The COGITO Study was supported by the Max Planck Society, including a grant from the Innovation Fund of the Max Planck Society (Grant M.FE.A.BILD0005); the Sofja Kovalevskaja Award (to Martin Lövdén) of the Alexander von Humboldt Foundation, donated by the German Federal Ministry for Education and Research; the German Research Foundation (Grant KFG 163); and the German Federal Ministry for Education and Research (Center for Advanced Imaging).

We thank the following persons for their important roles in conducting the COGITO Study: Annette Brose, Christian Chicherio, Gabi Faust, Katja Müller-Helle, Birgit Haack, Annette Rentz-Lühning, Werner Scholtysik, Oliver Wilhelm, and Julia Wolff, as well as a large number of highly committed student assistants.

Correspondence concerning this article should be addressed to Florian Schmiedek, Institute of Psychology, Humboldt-Universität Berlin, Unter den Linden 6, 10099 Berlin, Germany. E-mail: florian.schmiedek@psychologie.hu-berlin.de

### Empirical Findings

Across the adult lifespan, intraindividual trial-to-trial variability in RT performance increases for many tasks. This finding has been consistently reported across different research designs, such as comparisons between younger and older age groups (e.g., Anstey, Dear, Christensen, & Jorm, 2005), continuous age-related differences from younger to older adulthood (e.g., Li et al., 2004; Williams, Hulstsch, Strauss, Hunter, & Tannock, 2005), and longitudinal changes (Deary & Der, 2005; Fozard, Verduyssen, Reynolds, Hancock, & Quilter, 1994; MacDonald, Hulstsch, & Dixon, 2003).

Several studies investigated age differences in intraindividual RT variability on the basis of characteristics of RT distributions. These studies investigated either iSDs (e.g., Shammi, Bosman, & Stuss, 1998), quantiles of RT distributions (e.g., Salthouse, 1993), parameters of statistical distribution functions (e.g., the ex-Gaussian function; Spieler, Balota, & Faust, 1996; West, Murphy, Armilio, Craik, & Stuss, 2002), or parameters of theoretical process models that were fitted to the RT distributions (e.g., Ratcliff, Thapar, & McKoon, 2006a).

Regarding variability in cognitive performance at the day-to-day level, exceptionally little is known. A handful of studies have investigated variability in accuracy across testing sessions in memory performance (Hertzog, Dixon, & Hulstsch, 1992; Li, Aggen, Nesselroade, & Baltes, 2001), paper-and-pencil tests of cognitive abilities (Allaire & Marsiske, 2005), and perceptual-motor performance measures (Nesselroade & Salthouse, 2004). Only very few studies have examined daily variations in RT measures (e.g., Rabbitt, Osman, Moore, & Stollery, 2001; Ram, Rabbitt, Stollery, & Nesselroade, 2005; Sliwinski, Smyth, Hofer, & Stawski, 2006).

Antecedents of intraindividual variability in cognitive performance are likely to differ across time levels (e.g., seconds, minutes, hours, days, and years; cf. Martin & Hofer, 2004; Schaie, 1962) and across the course of skill acquisition (Li, Huxhold, & Schmiedek, 2004). With respect to lower level variability observed within shorter time ranges (e.g., single responses emitted in the second range), more basic cognitive mechanisms may play a prominent role, such as neuromodulatory processes regulating the efficiency of decision making (e.g., Li, Lindenberger, & Sikström, 2001; Ratcliff et al., 2006a). At the day-to-day level, external disturbances like stressful events (Sliwinski et al., 2006) and top-down influences like motivational aspects might come to the fore (Brose, Schmiedek, Lövdén, Molenaar, & Lindenberger, 2009). External disturbances and motivational factors may influence the amount of lower level variability by lowering the reliability of processing in the cognitive system or increasing the likelihood of attentional lapses. Conversely, lower level variability may statistically impinge upon the amount of observed day-to-day variability because daily performance is calculated as the mean of a limited number of (blocks of) trials, so that trial-to-trial variability is partly conserved in those means (cf. Rabbitt et al., 2001). Finally, longitudinal analyses suggest individual differences in developmental changes observed over years and decades share some of their etiology with variations in neural efficiency observed at the level of seconds (Lövdén, Li, Shing, & Lindenberger, 2007).

Processes related to skill acquisition may also transform the amount and conceptual meaning of intraindividual variability. Different cognitive abilities are known to differentially contribute

to overall performance (e.g., Ackerman & Cianciolo, 2000), and these shifts in contribution may also influence intraindividual variability. Intraindividual variability and changes therein may reflect changes in strategies (Siegler, 1994) or flag a transition from one system state to another (e.g., Bassano & Van Geert, 2007; van der Maas & Molenaar, 1992). In cognitive aging research, some findings indicate that practice may indeed influence the quality (e.g., Allaire & Marsiske, 2005) and quantity (e.g., Ram et al., 2005) of intraindividual variability.

### The Relation of Intraindividual Means and SDs

A common finding in most of the studies cited above is that age differences in RT variability, or longitudinal age-related and practice-related changes in RT variability, parallel those observed for RT means; RT variability increases with age and decreases with practice, but so does mean performance. Furthermore, iMs and iSDs mostly show strong positive correlations across individuals (e.g., Jensen, 1992). This has led researchers to call for and employ statistical controls for RT means when investigating individual and age differences in intraindividual RT variability. Two approaches are common. First, the *coefficient of variation* (CV; e.g., Guilford, 1956) is used to control for RT means by simply dividing iSDs by iMs. Second, regression analyses can be used to partial out RT means when investigating relations of iSDs to other variables (e.g., Salthouse & Berish, 2005). Both of these approaches are based on the assumption that the relation between iMs and iSDs is (a) linear and (b) invariant across time and individuals (cf. Wagenmakers, Grasman, & Molenaar, 2005). In this article, we demonstrate that both of these assumptions can be violated in empirical studies.

First, we summarize recent analyses of the relation of RT means and SDs. Because these are primarily based on variations across task conditions, we extend this discussion to variations across individuals and repeated measurement occasions. Second, we introduce *variance heterogeneity models* that allow direct investigation of the relation of iMs and iSDs and age-related differences therein. Third, we report a data set from the COGITO Study suitable for testing the invariant linearity assumption for younger and older adults and show that the assumption does not hold for this data set. Fourth, on the basis of this finding, we argue against the use of control techniques based on the invariant linearity assumption. Instead, we recommend the use of cognitive process models that express individual and age-based differences in the relation between central tendency and dispersion in terms of substantively interpretable parameter estimates. Using a simplified version of the diffusion model, we demonstrate with the data from the COGITO Study how such an approach can lead to insights about the mechanisms underlying age differences in the relation of iMs and iSDs. Fifth, because the relation between individual differences in iMs and iSDs is of interest in cognitive aging research contexts where both measures are used to predict certain criterion variables (e.g., dementia), we close with a discussion of the problem of differential reliabilities of iMs and iSDs, because this is another issue of practical relevance in cognitive aging research. In sum, although cognitive aging research is marked by a growing interest in understanding developing individuals as dynamic systems (Nesselroade, 1991), the field in general has typically treated intraindividual variability as complementary to

mean performance. In this article, we demonstrate and apply methodology that allows for answering questions about age differences and the processes underlying them by simultaneously modeling means and variances and their relations in both accuracies and latencies.

### *Is the Relation Between Means and Variances in Reaction Times Linear?*

Recently, the relation of mean RT and RT SDs across experimental task conditions has been investigated thoroughly by Wagenmakers and Brown (2007). They assembled three kinds of evidence to support a “law” of a linear relation of RT means and SDs. First, they summarized empirical findings showing good fit of linear functions that predict RT SDs with RT means across conditions of several experiments. Second, they showed that the relation between means and SDs of popular descriptive RT distributions, like the shifted Weibull or log-normal functions, is exactly or approximately linear if mapped out as a function of parameters related to task difficulty. Third, they demonstrated that important theoretical models for RT data, like Logan’s instance theory (Logan, 1988) or Ratcliff’s diffusion model (Ratcliff, 1978), predict linear relations of RT means and SDs.

Because the diffusion model has been successfully applied to aging data (e.g., Ratcliff et al., 2006a) and used in the empirical part of this article, it is introduced in some detail here. The diffusion model aims to explain data from two-choice RT experiments in a comprehensive way, including accuracy information as well as the shapes of RT distributions for correct and wrong responses. This is achieved by assuming a model for the decision process that involves several theoretically meaningful parameters. First, the quality of evidence accumulation during the decision process, called *drift rate*, is central. It describes how quickly information is accumulated in a random walk-like diffusion process that progresses from a starting point toward one of two response boundaries, one for correct and one for wrong responses. Large drift rates indicate fast accumulation of evidence; that is, an efficient decision process. The second central parameter of the model characterizes the carefulness of responding. This more strategic aspect of response behavior is implemented by differences in the distance between the response boundaries, called *boundary separation*. Wider boundary separation means more conservative responding because more evidence needs to be accumulated before a boundary is reached and a response is initiated. Still another parameter is *nondecision time*, combining peripheral sensory and motor aspects of the decision process. Advanced applications of the diffusion model also include additional parameters for variability of the central parameters across trials (see Ratcliff & Rouder, 1998).

As shown by Wagenmakers and Brown (2007), the parameter characterizing task difficulty (i.e., drift rate) can be used to explain the linear relation of means and SDs across task conditions. If only drift rate varies from one condition to the other, a linear relation of means and SDs results, because drift rate affects both the mean and the variance of the resulting RT distribution in a way that variance increases with the mean to the power of two. The resulting linear relation holds for the whole range of boundary separation values observed in empirical studies. Because variation in drift rates can also be used to characterize individual and age differences in the

efficiency of decision processes, this relation has the potential to explain the linearity of iMs and iSDs across people. If people not only differ in drift rates, however, but also in other parameters of the diffusion model, predictions for the relation of iMs and iSDs become more complex. The relation is not generally linear for differences in boundary separation (Wagenmakers et al., 2005) or nondecision time (Wagenmakers & Brown, 2007). Also, if people differ in conservatism of responding—an assumption not difficult to defend, particularly for different age groups—then the relation between iMs and iSDs does not have to be linear across individuals.

If the relation between iMs and iSDs across many repeated occasions is of interest, then practice-related changes in the diffusion model parameters also become an issue. Empirical evidence shows that all central diffusion model parameters can change with practice (Ratcliff, Thapar, & McKoon, 2006b). If nondecision times reduce with practice, as the motor aspects of responding are becoming more automatized and fluent, then the change influences iMs only but not iSDs, thereby distorting the linear relation between the two. As shown by Segalowitz and Segalowitz (1993), skill acquisition can indeed lead to changes in CVs; that is, to differential changes of iMs and iSDs.

To sum up, antecedents of intraindividual variability can be different depending on which time scale is analyzed, they can change with practice and with the selection of strategies, and they may differ across individuals and age groups. Such differences in the antecedents of intraindividual variability can also lead to differences in the relation of iMs and iSDs. Hence, the relation between RT means and SDs cannot safely be assumed to be linear and invariant across individuals and age groups and therefore needs to be investigated empirically.

### *Modeling the Relation of Means and Variances in Reaction Times*

Mean performance can be calculated for each single time unit at a given level of time. However, iSDs need to be calculated across several of the time units at the time level of analysis. For example, at the time level of daily occasions, mean performance can be calculated separately for each day. However, to compute iSDs using conventional methods, performance of several days has to be binned before iSDs can be calculated. This reduces the time resolution for iSDs and potentially also influences their interpretation.

In the context of multilevel or mixed models (e.g., Snijders & Bosker, 1999), elegant and efficient methods for overcoming this problem and investigating relations between iM and iSD have been developed (e.g., Hoffman, 2007). These models (sometimes called *location-scale models*, *dispersion models*, or *models with heterogeneous variances*) do not require binning of the data because they model the *expected variance* at any time point. Although more than one time point is needed to calculate an SD, statistical expectancies for the amount of variance can be formulated for a single time point, and those expectancies can be further modeled as functions of mean performance or other predictor variables.

Two classes of models of this kind are *power of means* (POM) models and log-linear variance heterogeneity models. POM models assume that the expected variance at a certain time point is a power function of the mean at this time point, for example,

$$\hat{\sigma}_{ij}^2 = \phi_i |\hat{X}_{ij}|^{\theta_i}, \quad (1)$$

where  $\hat{\sigma}_{ij}^2$  is the expected variance of person  $i$  on occasion  $j$ ,  $\hat{X}_{ij}$  is the expected mean performance for person  $i$  on occasion  $j$  (which might result from any linear or nonlinear model for mean performance),  $\theta_i$  is the exponent of the power function for person  $i$ , and  $\phi_i$  is an additional (person-specific) scaling factor for the relation of mean performance and variance. In the case of a common linear relation of iMs and iSDs, the estimated exponent would need to be 2 for all persons. In addition, the scaling factor would have to be equal across persons to assure comparable steepness of the linear function relating iMs and iSDs.

The alternative log-linear models (Harvey, 1976) assume an exponential relation of means and variance; for example,

$$\hat{\sigma}_{ij}^2 = \phi e^{(\beta_0 + \beta_1 \hat{X}_{ij})}. \quad (2)$$

Here, the expected variance is an exponential function of a linear term, which in addition to expected mean performance might also include other additional predictor variables. These models therefore allow investigating different models for mean performance and for variability around it, thereby providing a versatile tool for investigating many theoretical questions in research on intraindividual variability. Both the POM and the log-linear models can be implemented in simplified versions that include only two parameters for all persons, one for the nonlinear relations between means and variances, and a scaling factor. In the POM model, the power exponent captures the nonlinearity of the relation of iMs and iSDs, whereas the scaling factor defines the steepness of the slope. Mean RTs of zero are associated with zero variance. Similarly, the coefficient  $\beta_1$  in the exponent term (in the following also simply called exponent) of the log-linear model captures the acceleration of the functional relation between iMs and iSDs. The coefficient  $\beta_0$  can also be interpreted as a scaling factor because  $\phi e^{(\beta_0 + \beta_1 \hat{X}_{ij})} = \phi e^{(\beta_0)} e^{(\beta_1 \hat{X}_{ij})}$ . This parameter can therefore be used to capture individual and group differences in the steepness of the functions. It also influences the predicted amount of intraindividual variance at mean RTs of zero. Because mean RTs virtually cannot be zero, and if they were, variance would have to be zero as well, the estimated intercept of  $\phi e^{(\beta_0)}$  is not directly interpretable. However, the property of the log-linear model of having an intercept allows for larger amounts of baseline variance at some empirical minimum of observed mean RTs than is possible for the POM model.

### The COGITO Study

In the empirical part of this article, longitudinal data from the COGITO Study are used. The central aim of the COGITO Study was to investigate different levels of intraindividual variability of cognitive functioning from a multivariate perspective (Lindenberg, Li, Lövdén, & Schmiedek, 2007). To this end, samples of 101 younger and 103 older adults came to lab rooms in central Berlin, Germany, to work on a comprehensive battery of computerized cognitive ability tests for an average of 100 daily sessions. The cognitive test battery comprised a total of 12 tests for the ability constructs of perceptual speed, episodic memory, and working memory, each operationally defined with tasks from verbal, numerical, and figural-spatial content domains. To be able to keep task difficulty constant across the 100 daily sessions without

producing floor or ceiling effects, presentation times of the episodic and working memory tasks were set to fixed individualized values on the basis of time-accuracy functions fitted to pretest performance data. For the present investigation, we are using one of the working memory tasks, a spatial version of the  $n$ -back paradigm (e.g., Cohen et al., 1997), which combines the need for continuous updating of several items to be kept in working memory with the requirement of rapid two-choice decisions for each item, thereby producing RT information as one dependent variable. We use this data to (a) report descriptive findings on age differences in intraindividual RT variability observed at the day-to-day level; (b) investigate, with POM and log-linear models, whether the parameters describing the relation of iMs and iSDs differ across age groups and across individuals within age groups; and (c) query whether diffusion model parameters estimated separately for each participant and occasion can provide explanations for the predicted age differences in the relation of iMs and iSDs.

## Method

### Participants and Procedure

Participants were recruited through newspaper advertisements, word-of-mouth recommendation, and flyers distributed in university buildings, community organizations, and local stores. The advertisements addressed people interested in practicing cognitive tasks for 4–6 days a week for a period of about 6 months. Allowances were mentioned, but no detail was given about the amount. It took several steps to get included in the study. First, in telephone interviews, interested persons were given information about the study, and we checked whether requirements for study participation, in particular time investment, could be met. Potential candidates for participation were then called back and invited to join a 1-hr “warm-up” group session to get more information about the study. General aims of the study were explained and detailed information on incentives was given. The digit-symbol substitution test and a questionnaire on sociodemographic variables were administered. Individuals could assign themselves to the study after the end of this session. Participants underwent 10 days of pretests held in group sessions (2.0–2.5 hr); they included a large number of self-report questionnaires, instruction to and behavioral testing with tasks included in the daily protocol, and a number of covariates and transfer tasks. During the longitudinal phase, participants scheduled daily sessions (1.0–1.5 hr) on an individual basis on up to 6 weekdays (including Saturdays) per week. Participants worked on the tasks individually in rooms with three to six work places. At the end of each session, participants got feedback on their performance on all tasks, including average accuracies and RTs. They could get printouts of these results to take home. At posttest, another 10 group sessions (1.5–2.0 hr) were conducted with repeated administration of the pretest cognitive tasks and additional self-report measures.

The final sample included 101 younger (51.5% women; age: 20–31,  $M = 25.6$ ,  $SD = 2.7$ ) and 103 older (49.5% women; age: 65–80,  $M = 71.3$ ,  $SD = 4.1$ ) adults who did an average of 101 sessions during an average total time of 192 days.

### The Spatial 3-Back Task

In each daily session, participants worked on four trials of the spatial 3-back task. In each trial, a sequence of 39 black dots



appeared at varying locations in a  $4 \times 4$  grid. Participants had to respond and indicate whether each dot was in the same position as the dot three steps earlier in the sequence (green key on a button box) or not (red key). Participants were instructed to respond as quickly and accurately as possible. Dots appeared at random locations with the constraints that (a) 12 items were targets; (b) dots did not appear in the same location at consecutive steps; (c) exactly three items each were 2-, 4-, 5-, or 6-back lures (i.e., items that appeared in the same position as the items 2-, 4-, 5-, or 6 steps earlier). The first three items in each sequence were not used in the analyses because they could never be targets. Presentation rate for the dots was individually adjusted on the basis of pretest performance. This was done by keeping the presentation time of the dots constant at 500 ms but varying the interstimulus intervals (ISIs). Eighty-two younger

and 27 older adults were assigned to the fastest (ISI 500 ms), eight younger and 16 older adults to the intermediate (ISI 1,500 ms), and 11 younger and 60 older to the slowest (ISI 2,500 ms) possible presentation rate.

Before beginning the longitudinal phase, participants had already practiced the spatial 3-back task for 12 blocks for each of four different presentation rates (500 ms, 1,500 ms, 2,500 ms, and 3,500 ms) in one pretest session.

#### Data Analysis Methods

Given that the goal of the analyses was to investigate intraindividual variability at the day-to-day level, it was necessary to account for slower trends in the data. Figure 1 shows average trends of RT iMs, RT iSDs, and accuracies, indicating that for

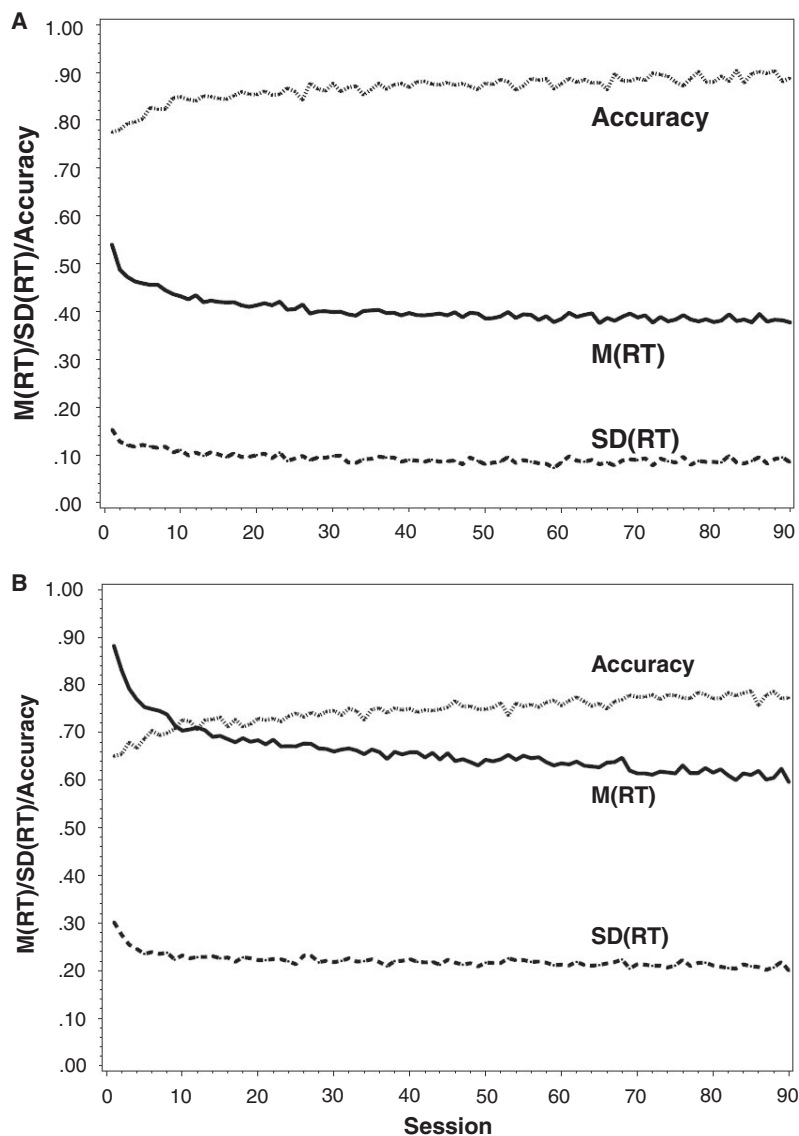
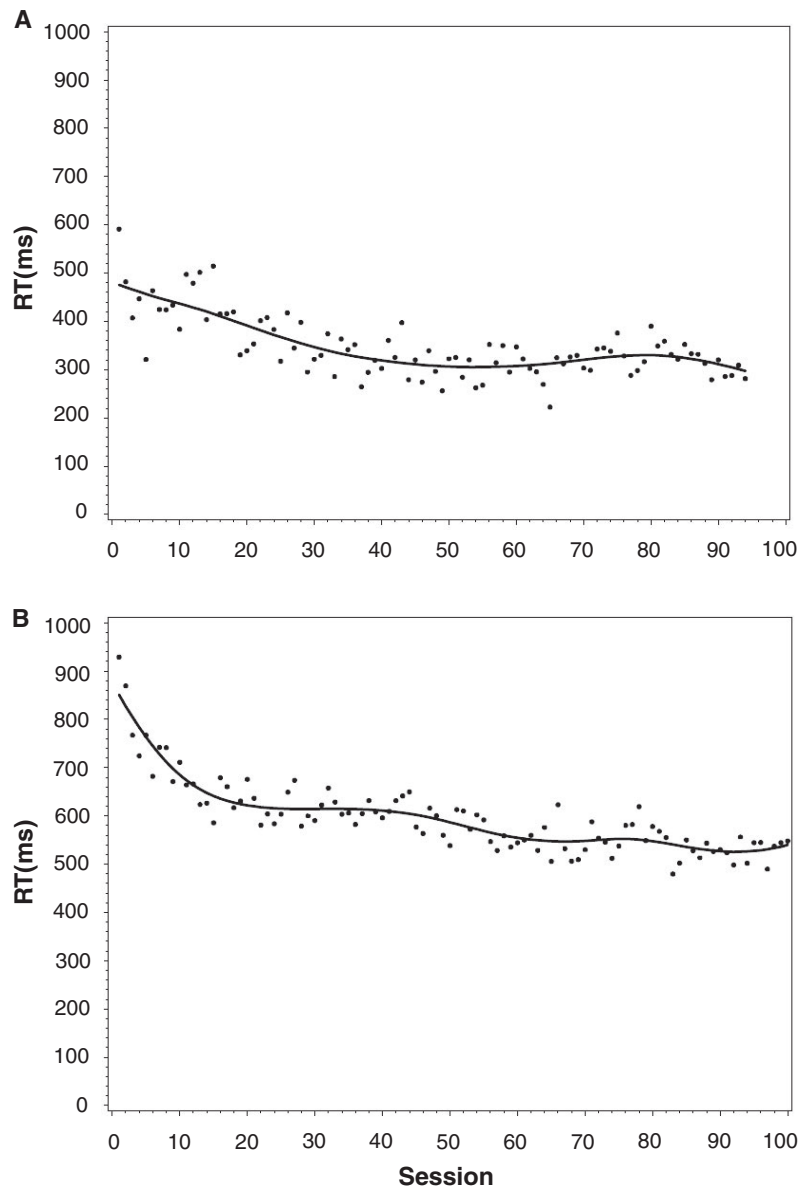


Figure 1. Mean trends of reaction time (RT) intraindividual means (iMs; in seconds; solid lines), RT intraindividual standard deviations (iSDs; in seconds; broken lines), and accuracies (proportion correct; dotted lines). A: Younger adults. B: Older adults.

both age groups, performance improved across the testing sessions. Individual trends for several individuals did not show continuous improvements and therefore could not be captured sufficiently well with theoretical learning curves, such as exponential functions, but exhibited more complex patterns of intraindividual changes. Therefore, we decided to describe mean performance changes with penalized radial spline smoothing functions as implemented in SAS PROC GLIMMIX (SAS Institute, 2006). With this semiparametric method, trends are not fitted separately to each individual's time series but to all individuals simultaneously, using a mixed model approach, with individual differences in the functions captured by random

effect parameters (Ruppert, Wand, & Carroll, 2003). Predicted values can be created using best linear unbiased prediction methods (see Ruppert et al., 2003; and Appendix A). Visual inspection of these predicted trends together with the observed data indicated that even the more deviant patterns of intraindividual change were captured sufficiently well with this approach. Examples of observed mean RTs and trends fitted with this method for one younger and one older adult are shown in Figure 2, whereas fitted trends for the whole samples of younger and older adults are shown in Figure 3. As a comparison approach, exponential learning curve functions were fitted separately to each individual's time series and residuals ana-



*Figure 2.* Examples of individual trends fitted with the penalized radial spline smoothing method for one younger and one older adult with individual exponents of the power-of-means (POM) function close to the group mean. Dots = observed mean reaction times (RTs) for each daily session; line = fitted trend. A: One younger participant with POM exponent = 3.69. B: One older participant with POM exponent = 2.45.

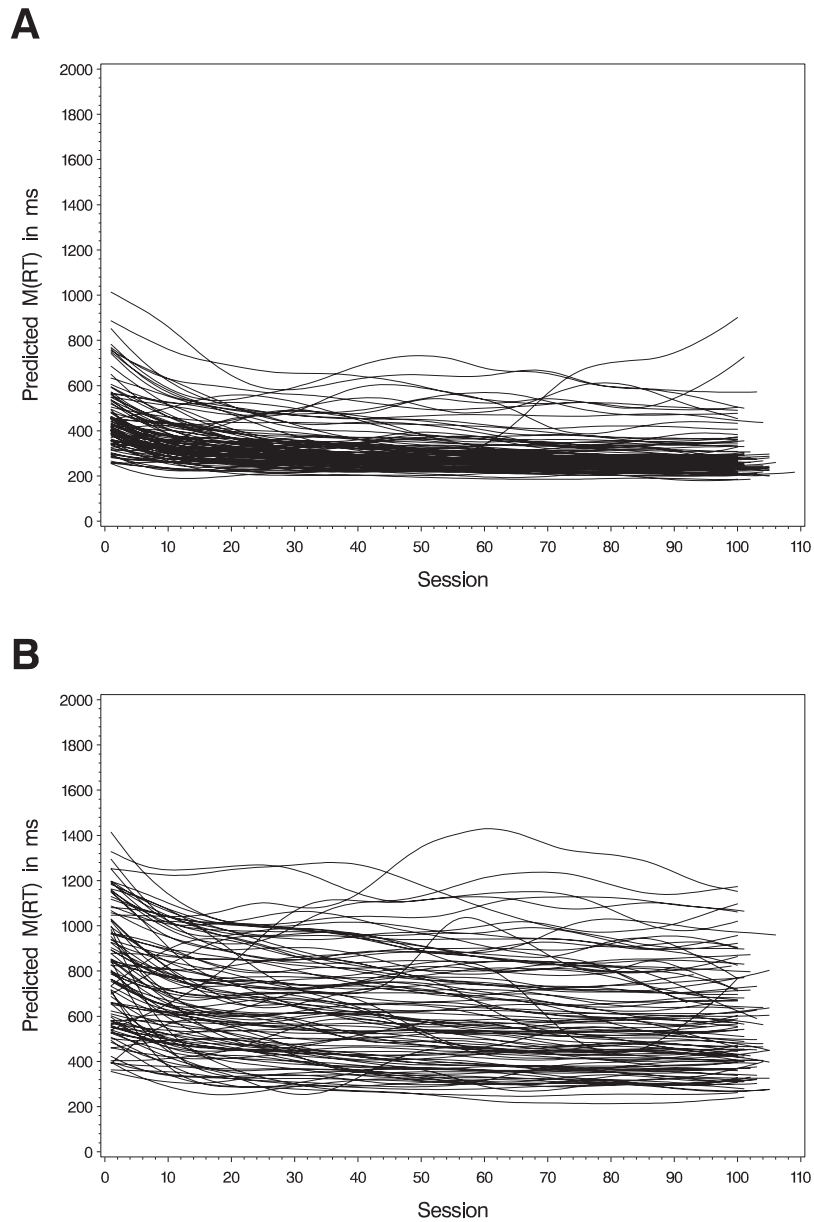


Figure 3. Individual reaction time (RT) trends as predicted by semiparametric spline smoothing functions. A: Younger adults. B: Older adults.

lyzed in the same way as the residuals from spline smoothing (see Appendix B for the model and results).

POM and log-linear variance heterogeneity models were fitted with SAS PROC MIXED (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2007) with the LOCAL = EXP() function for the log-linear and the LOCAL = POM() function for the POM models, using the predicted values from the fitted spline and exponential functions as the model of mean performance. The corresponding residuals that resulted from subtracting the fitted functions from the raw data were plugged into the models as dependent variables. The alpha level for statistical significance was set to  $p = .05$ .

## Results

Results are organized in three sections. First, we report results from the mean trend fitting and descriptive findings on age differences of intraindividual variability around these mean trends. Second, we present results from fitting the two variance-heterogeneity models, the POM and log-linear models, and age differences in the parameters describing the relation between iMs and iSDs. Third, the EZ diffusion model (Wagenmakers, van der Maas, & Grasman, 2007) is fitted to individuals' response RT data to arrive at estimates of diffusion model parameters for each participant and occasion. These parameters are then used for

interpreting the observed age differences in the relation of iMs and iSDs.

*Mean Trends and Day-to-Day Variability*

Mean trends as captured by the penalized spline functions are shown in Figure 3. ISDs were calculated by subtracting these predicted values from observed mean RTs for each session and calculating individual SDs for the residuals. On average, these iSDs were larger for older adults,  $M(iSD) = 44$  ms,  $SD(iSD) = 29$  ms, than for younger adults,  $M(iSD) = 29$  ms,  $SD(iSD) = 26$  ms. This age group difference in intraindividual RT variability was significant,  $t(202) = 3.84$ ,  $p < .05$ . When individual trends were fitted with exponential functions, a comparable pattern emerged. iSDs based on residuals from such functions were also significantly larger for older than for younger adults: younger adults,  $M(iSD) = 34$  ms,  $SD(iSD) = 30$  ms; older adults,  $M(iSD) = 54$  ms,  $SD(iSD) = 36$  ms;  $t(202) = 4.32$ ,  $p < .05$ .

*Variance Heterogeneity Models*

*POM model: Group fits.* Results for fitting the POM model to the group data are shown in Table 1 and Figure 4. The baseline model, which assumes just one overall function, common for both age groups, for the relation of iMs and iSDs, led to an estimated power exponent of 2.27. This implies a close-to-linear relation of

iMs and iSDs (Figure 4A). If we had adopted this model, results would have been interpreted as nicely supporting the purported law of a linear relation of iMs and iSDs. However, when parameters of the POM model were allowed to vary across age groups (Figure 4B–C), it became evident that model fit improved considerably and reliably. Specifically, when the parameters were allowed to differ across groups, the minus 2 log-likelihood ( $-2LL$ ) decreased by 2,086 units with just two additional parameters,  $\Delta\chi^2(2) = 2,086$ ,  $p < .05$ . The resulting functions for the two groups were radically different (Figure 4C). Whereas the function for the older group continued to be fairly linear but shallower than for the baseline model, the function for the younger group had an exponent much larger than two and therefore deviated from the assumption that the relation between iMs and iSDs is governed by a linear function.

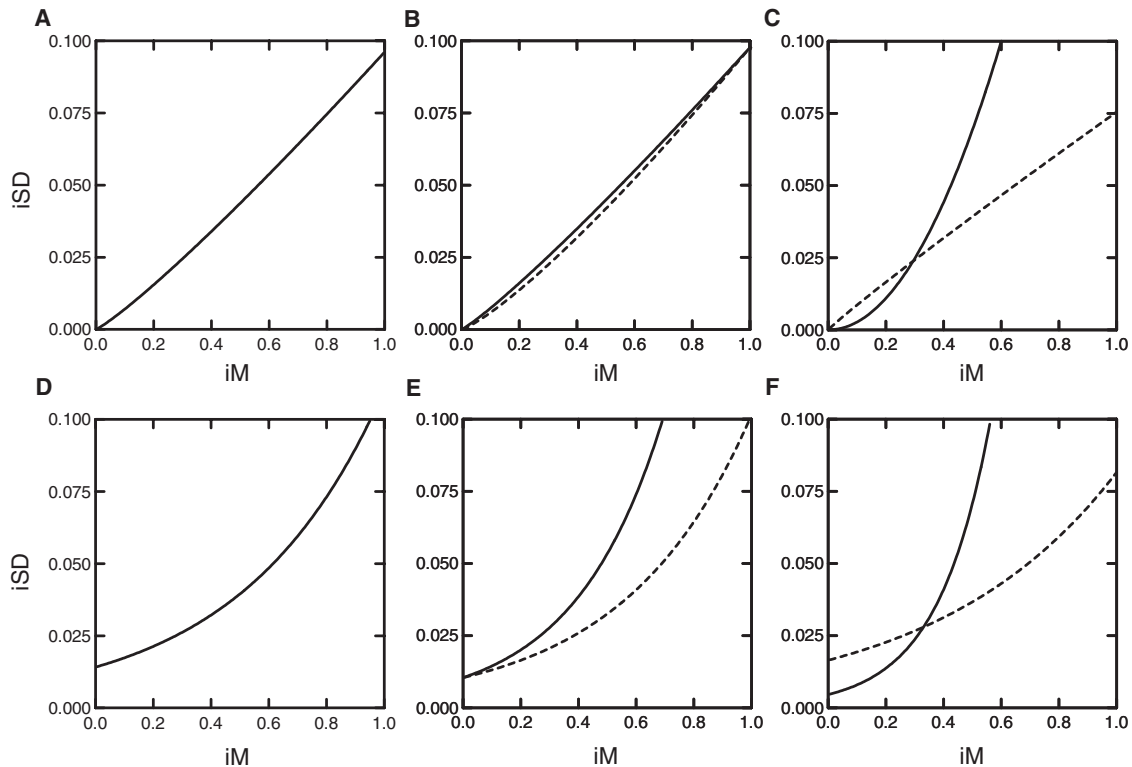
*POM model: Individual fits.* The POM model was also fitted separately to each individual’s time series of predicted iMs and iSDs. This was possible without estimation problems for all persons. The cumulated  $-2LL$  over these individual fits was  $-86,462$ . The difference in  $-2LL$  to the model with group specific parameters was 7,049, which is highly significant given the difference in number of parameters (404). Therefore, individuals differed considerably in their relations between iMs and iSDs. As apparent in Figure 5, these functional relations were shallower and much more heterogeneous for the older group. Age group differences were

Table 1  
Results From Fitting POM and Log-Linear Variance Heterogeneity Models

Variable	POM	Log-linear
	Baseline	
Exponent	2.27	4.11
Scaling factor	0.00924	0.00020
Number of parameters	2	2
Model fit ( $-2LL$ )	$-77,326$	$-75,991$
	Different exponents for age groups	
Exponent (younger)	2.25	6.53
Exponent (older)	2.45	4.55
Scaling factor	0.00954	0.00011
Number of parameters	3	3
Model fit ( $-2LL$ )	$-77,398$	$-77,651$
	Different exponents and scaling factors for age groups	
Exponent (younger)	4.03	10.95
Exponent (older)	1.89	3.20
Scaling factor (younger)	0.07831	0.00002
Scaling factor (older)	0.00570	0.00027
Number of parameters	4	4
Model fit ( $-2LL$ )	$-79,413$	$-79,308$
	Different exponents and scaling factors for individuals	
	$M$ ( $SD$ )	$M$ ( $SD$ )
Exponent (younger)	3.67 (2.59)	11.49 (8.96)
Exponent (older)	2.59 (3.71)	4.35 (6.91)
Scaling factor (younger)	24.2739 (212.0834)	0.00066 (0.0027)
Scaling factor (older)	0.4106 (3.0380)	0.04329 (0.2740)
Number of parameters	408	408
Model fit ( $-2LL$ )	$-86,462$	$-88,082$

Note. POM = power of means.





*Figure 4.* Estimated functions relating intraindividual reaction time (RT) means (iM; in seconds) to intraindividual RT standard deviations (iSD; in seconds) based on: (a) power-of-mean variance heterogeneity models (Panels A, B, and C); (b) log-linear variance heterogeneity models (Panels D, E, and F). A and D: Baseline models. B and E: Models with age-group-specific exponents. C and F: Models with age-group-specific exponents and scaling factors. Solid lines = younger adults; broken lines = older adults. Functions reach beyond the observed data range, as the fastest RT means were 179 ms for younger and 213 ms for older adults.

significant for exponents,  $t(202) = 2.41$ ,  $p < .05$ , but not for scaling factors,  $t(202) = 1.14$ .

*Log-linear model: Group fits.* Results for fitting the log-linear variance heterogeneity model parallel those for the POM model: Allowing the exponents and scaling factors to differ between age groups strongly improved fit, and older participants on average showed a shallower relation between iSDs and iMs than younger participants (see Table 1 and Figure 4D–F). The two models differed in the sense that the log-linear functions were more strongly bent and suggested greater baseline variability at minimum levels of mean RT for the older group. Keep in mind, however, that RTs around 200 ms represent an extrapolation beyond observed data for most of the older participants.

*Log-linear model: Individual fits.* The log-linear model was also fitted separately to each individual's data. Again, this was possible without estimation problems for each person. The cumulated  $-2LL$  was  $-88,082$ , a highly significant improvement over the group model ( $-2LL$  difference =  $8,774$ ;  $df = 404$ ). Plots of the individual functions, which are not shown here, indicated that the individual fits of the log-linear model were very similar to the fits obtained with the POM model. Again, age group differences were significant for exponents,  $t(202) = 6.38$ ,  $p < .05$ , but not for scaling factors,  $t(202) = -1.56$ .

In addition to models based on residuals from spline-smoothing trends, all of the analyses reported in this section were also carried

out for residuals that were based on individually fitted exponential functions. In this manner, trends that deviate from the theory-consistent exponential and operate on a slower frequency than days are preserved in the residuals. Results for these more inclusive definitions of intraindividual variability were very similar to the findings reported above (see Appendix B).

### *Diffusion Model*

To explore potential reasons for the large differences in the relations between iMs and iSDs, both between age groups and across individuals within age groups, the response time data from each participant and each daily session were parameterized in terms of the diffusion model. The main rationale for conducting these analyses derived from the consideration that the relative contribution of reductions in nondecision times to the overall speedup of responding with practice was greater in older adults than in younger adults. Nondecision time refers to sensory and motor aspects of responding and does influence iMs but not iSDs, which means that reductions in nondecision times could equally lead to shallower iM–iSD functions. In sum, then, the objective of the diffusion model analyses was to test the hypotheses for the observed age differences in the relations between iM and iSD that older adults showed greater reductions in nondecision times than younger adults.

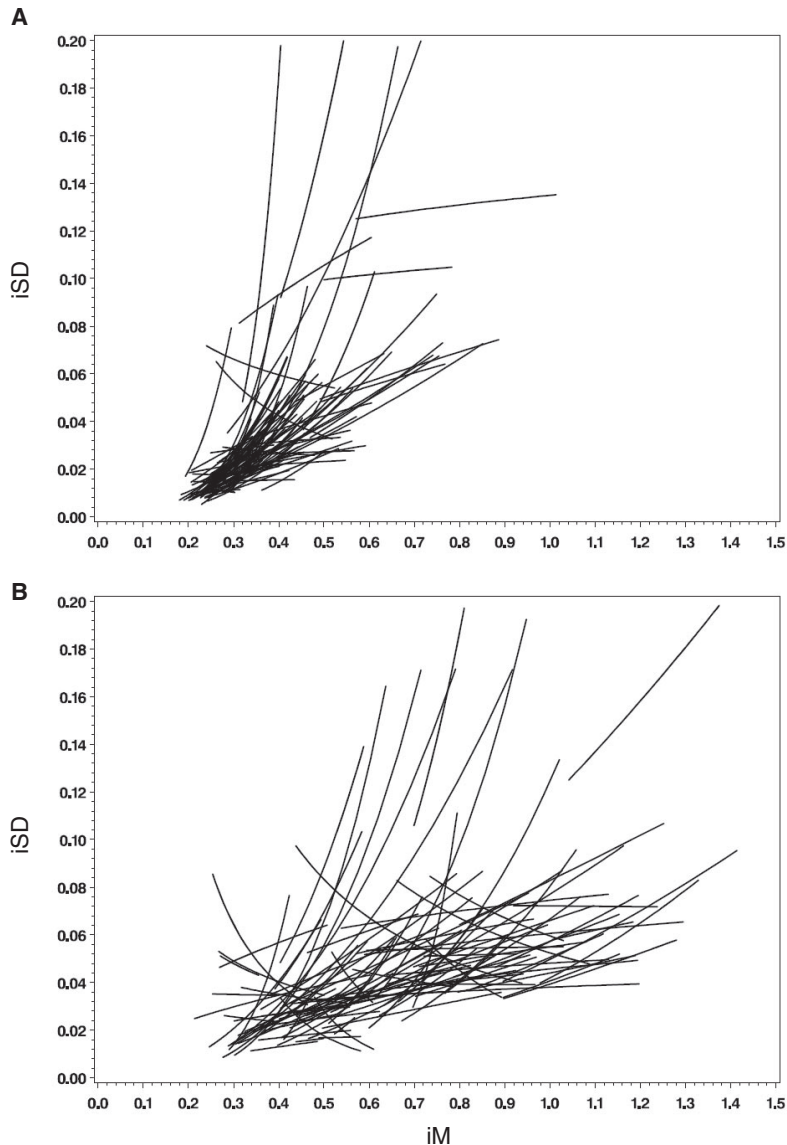


Figure 5. Functions relating intraindividual reaction time (RT) means (iM; in seconds) to intraindividual RT standard deviations (iSD; in seconds) based on the power-of-means variance heterogeneity model fitted separately to each individual's time series data. A: Younger adults. B: Older adults.

Estimation of the full diffusion model with all variability parameters was not possible because the number of responses per daily session (i.e.,  $4 \times 36$  responses) was too small. Therefore, we used the simplified EZ approach to diffusion modeling (Wagenmakers et al., 2007). The EZ approach has been successfully applied to data sets with a small number of RTs (e.g., Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). EZ calculates the three central parameters of the diffusion model—drift rate ( $v$ ), boundary separation ( $a$ ), and nondecision time ( $T_{er}$ )—directly from average accuracy and the mean and variance of correct RTs in closed form. RTs shorter than 300 ms were excluded before parameter calculation (cf. Ratcliff, Thapar, & McKoon, 2004). Figure 6 shows average trends for younger (Panel A) and older adults (Panel B). Apparently, younger adults improved more on

drift rates whereas older adults showed a stronger reduction in nondecision time. Boundary separation was relatively stable for both groups.

Within-person linear regressions of nondecision time on session number were calculated to estimate the improvements of nondecision time across the 100 days of training to test the hypothesis that older adults might reduce their nondecision times more strongly than younger adults. The average regression weight of nondecision time on session number, indicating reductions in nondecision time, was larger for older adults,  $\beta = -.0010$ ,  $SE = .0001$ , than for younger adults,  $\beta = -.0004$ ,  $SE = .00005$ . The difference between the two regression weights was reliable,  $t(202) = 5.55$ ,  $p < .05$ . The observed age group difference in the exponent of the POM model could be partly explained by individ-

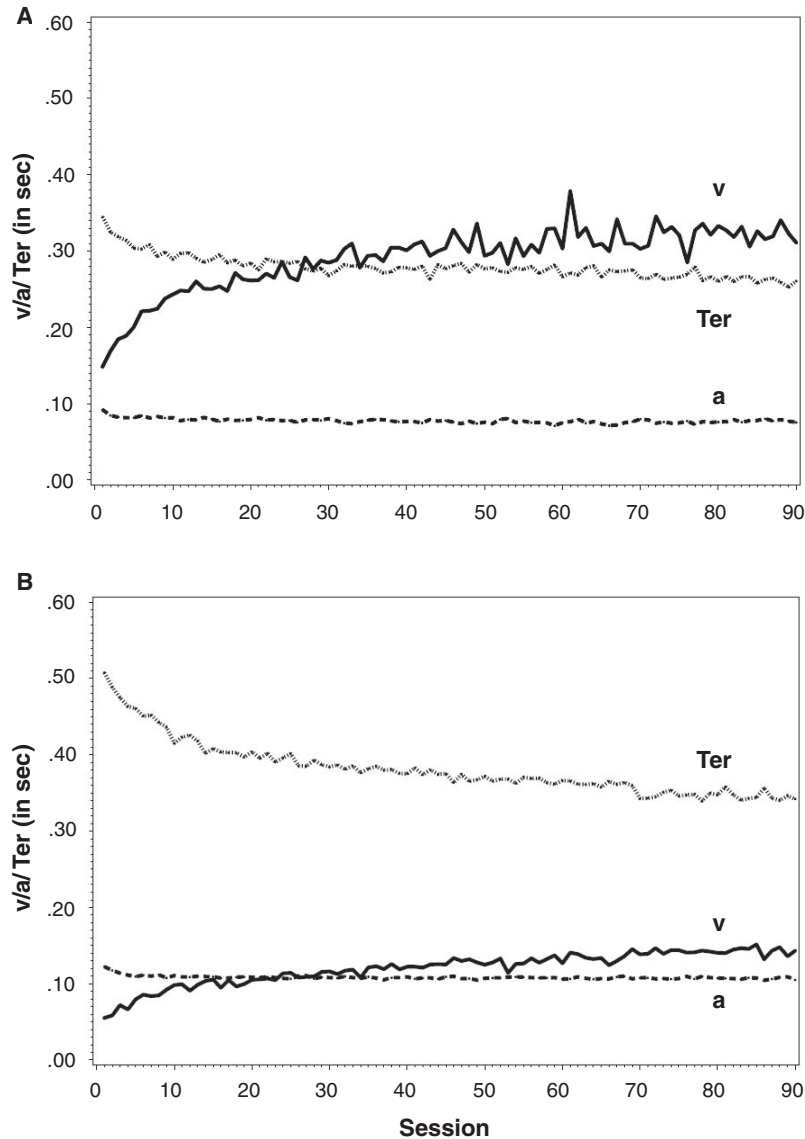


Figure 6. Mean trends of diffusion model parameters estimated with the EZ diffusion model.  $v$  = drift rate (solid lines);  $a$  = boundary separation (broken lines);  $Ter$  = nondecision time (dotted lines). A: Younger adults. B: Older adults.

ual differences in the reductions of nondecision time. When entering both age group and the regression weights from the within-person regressions of nondecision time on session number in a between-person regression analysis with the POM exponent as dependent variable, the effect of reductions in nondecision time was significant,  $F(1, 201) = 4.00, p < .05$ , reducing the age group difference to a nonsignificant value,  $F(1, 201) = 2.35$ . It seems that larger reductions in nondecision time in the older group are at least contributing to the explanation of observed individual differences in the relation between iMs and iSDs. Because age differences in the relations between iMs and iSDs could partly be explained by age differences in reductions of  $Ter$ , we investigated whether conducting our analyses (spline smoothing and fitting of variance heterogeneity models) on mean *decision time* (i.e., mean

RT minus  $Ter$ ) rather than mean RT would reduce the observed age differences in the POM exponents. This turned out to be the case. Functions were much shallower and more similar across the two age groups.<sup>1</sup>

## Discussion

In this article, we explore individual differences in the relation between central tendency and dispersion of working-memory related reaction times, with an emphasis on group differences between younger and older adults. For this purpose, we fitted two

<sup>1</sup> We thank Eric-Jan Wagenmakers for suggesting this kind of analysis.

different multilevel variance heterogeneity models, POM and log-linear, to estimate age-group differences as well as individual differences within age group in the relations between iM and iSD. At first sight, the fitted overall POM function looked like another demonstration of a linear relation between iMs and iSDs, as its exponent was fairly close to the value of two. However, model fits increased dramatically when parameter heterogeneity between age groups and among individuals within age groups was permitted. By analogy, ignoring this heterogeneity and reverting to the group model with a “one-size-fits-all” exponent is as prohibited as assuming a material-invariant coefficient of linear thermal expansion  $\alpha$  for materials as different as diamond ( $\alpha = 1$ ), gold ( $\alpha = 14$ ), and rubber ( $\alpha = 77$ ). Thus, researchers investigating age-group and individual differences in iSDs who attempt to control for individual differences in iMs by using statistical control procedures that assume an invariant linear relation between iM and iSD risk biased and misleading results. Specifically, the use of the CV as well as the use of linear statistical control techniques such as partial correlations or multiple regression analysis is dubious if heterogeneity exists in the relation between iMs and iSDs. Hence, we argue that findings from past research on age differences in iSD that have been obtained with methods positing invariant linearity between iM and iSD need to be interpreted with caution.

To overcome the impasse created by the apparent inappropriateness of standard linear control techniques, we demonstrated the utility and versatility of two complementary approaches for modeling heterogeneity in the relations between iMs and iSDs in different groups and different individuals. The first approach is to explicitly model the relation between iMs and iSDs with variance heterogeneity models of the kind used here. Whereas POM models allow us to directly investigate how much the data deviate from a linear function, as the linear relation between iMs and iSDs is contained in the POM model as a special case, log-linear models as implemented in PROC MIXED have the potential to explore and test models predicting the amount of intraindividual variability with time-invariant and time-varying predictor variables in parallel to models for mean performance. It is therefore possible to model different antecedents for mean performance and intraindividual variability around it (cf. Hoffman, 2007)—a highly advantageous option for many interesting research questions. Taking these ideas one step further, recent methodological developments even allow us to model interindividual differences in the amount of intraindividual variability as random effects (in so-called *random scale models*; Hedeker, Mermelstein, & Demirtas, 2008).

The second approach proposed here is to attempt to explain the relation between iMs and iSDs with theoretical process models. We showed how one prominent process model for two-choice decision tasks, the diffusion model, can be used to better understand observed age differences in the relations between iMs and iSDs. In short, we found that relative to younger adults, older adults' improvements in overall RT were more strongly determined by improvements in sensory and motor aspects of responding, as estimated by the nondecision time parameter of the diffusion model. Changes in nondecision time exclusively influence iMs and thereby lower the slope of the iM–iSD function—in this case, selectively for older adults. Performance on the *n*-back task, with its requirement of constantly updating the contents of working memory, likely involves a complex set of processes that potentially includes binding, inhibition, and reliance on familiarity

information (Schmiedek, Li, & Lindenberger, 2009). Practice-induced improvements on this task can be considerable (Li, Schmiedek, Huxhold, Röcke, Smith, & Lindenberger, 2008). As the results reported here indicate, the mechanisms underlying the observed improvements in accuracy and RT could involve changes in several processes, including improvements in nondecision time. However, most important for improvements in both accuracy and RT are improvements in drift rate. These may originate from stronger representations of the current target position that the current stimulus needs to be compared with, which in turn might result from more reliable updating operations, stronger bindings of spatial to temporal positions, better unbinding or inhibition of no longer relevant stimuli, or combinations of the above.

In this article, we question the use of statistical control techniques that assume invariance of relations between iM and iSD across age groups and individuals and urge researchers to explicitly model these relations at the levels of age groups and individuals. In doing so, we follow recent calls to test the *ergodicity assumption* (Molenaar, 2004), which states that the structure of between-person differences matches the structures of within-person variability. We were able to show that the relations between iMs and iSDs of working-memory related reaction times are non-ergodic. As the within-person relations between iMs and iSDs differed reliably and widely across people, the roughly linear relation of between-person differences in means and standard deviations tells us close to nothing about the functional relations between means and standard deviations within a given person. Therefore, any standardization or calibration of iSDs in terms of their between-person relation to iMs will result in indices that fail to describe, explain, or predict within-person processes, such as performance differences across experimental conditions or practice-induced improvements.

Hence, the marked heterogeneity in the functional relations between iMs and iSDs suggests that a surrogate approach, in which between-person variability is taken as a proxy for within-person variability, is not tenable. Instead, process models of RT variability need to be tested at the within-person level. One could be tempted to argue that the divergence of between-person and within-person findings is mainly due to practice-induced changes of the underlying processes, so that studies investigating individual and age differences in mean and variability measures with much smaller amounts of practice (e.g., within a single session) might be less affected by these issues. This is only the case, however, to the degree that individual differences without experimental practice are not also determined by pre-experimental individual differences affecting relevant processing parameters. It could well be that numerically identical CV are observed in two different individuals, but that the two CV come about by entirely different constellations of drift rate, boundary separation, and nondecision time, even before experimental practice affects them, rendering this value uninformative about underlying mechanisms.

We hasten to add that the accurate representation of within-person processes is only one of several goals of behavioral research. In clinical settings, for instance, the prediction of group membership (e.g., dementia status) may be more important. Here, issues of explanation and prediction primarily arise at the between-person level, and the modeling of within-person relations is less relevant. In these instances, it may make sense to control for individual differences in iMs to probe the incremental validity of

iSDs, assuming that the two show a linear and positive correlation at the between-person level.

### Reliability of iMs and iSDs

If testing incremental validity of iSD over iM at the between-person level is the goal, one needs to pay close attention to another methodological issue: the reliability of iMs and iSDs as measures of interindividual differences (Schmiedek, 2006). When iMs and iSDs that were computed on the basis of the same given number of occasions (or trials, or items) compete against each other in a regression analysis predicting a criterion variable of interest, the relative size of the regression weights for the two predictor variables as well as the corresponding amounts of incremental variance explained will be influenced by the relative reliabilities of iMs and iSDs. If one of the two is more reliable, then this does put the competitor at a disadvantage.

According to classical test theory, reliability is the ratio of true to observed variance, and observed variance is the sum of true and error variance. Thus, iMs and iSDs will have the same reliability (for a given number of occasions) if the ratio of error to true variance is the same for both. Whereas the relative amounts of true variance have to be determined empirically, the relative amounts of error variance can be derived analytically. Statistical textbooks tell us that the standard error of a standard deviation is .707 times as large as the standard error of the mean (e.g., McNemar, 1962, p. 78). This implies that for a given number of occasions, the error variance of an iSD will only be half ( $.707^2 = .50$ ) as large as the error variance of an iM. Whereas this general formula does not apply to very small numbers of occasions, it is easy to show with simulations that the error variance of a standard deviation is always smaller than that of the mean (see Figure 7). Regarding the amount of error variance due to sampling a few occasions out of a population of occasions, iSDs are therefore, surprisingly perhaps, at an advantage in comparison to iMs.<sup>2</sup>

However, reliability does also depend on the relative amount of true variance, which cannot be determined analytically. Thus, the question of whether iMs or iSDs have comparable reliabilities

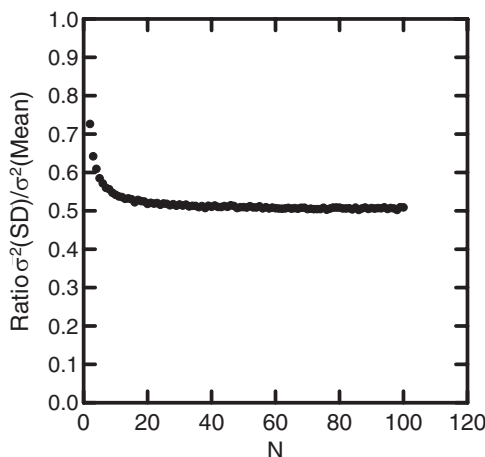


Figure 7. Ratio of the error variance for estimating a standard deviation to the error variance for estimating a mean given a certain number of observations ( $N$ ).

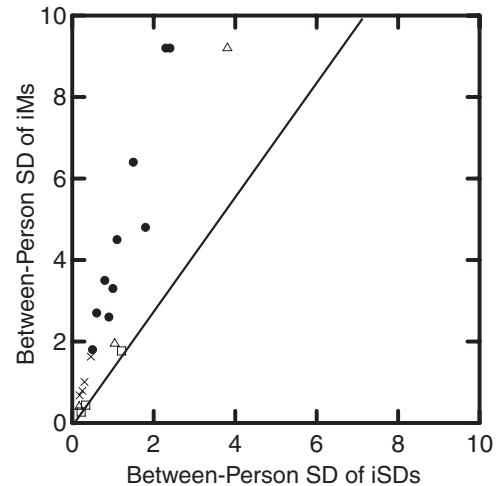


Figure 8. Empirical ratios of between-person standard deviations of intraindividual means (iMs) versus intraindividual standard deviations (iSDs) from published studies (circles, Salthouse, Nesselroade, & Berish, 2006; triangles, Lecerf, Ghisletta, & Jouffray, 2004; squares, Salthouse & Berish, 2005; crosses, Nesselroade & Salthouse, 2004). If the ratio is larger than the square root of 2 (area above diagonal line), the reliability of intraindividual means is larger than the reliability of intraindividual standard deviations.

boils down to an empirical question. It is easy to derive that if the observed between-person SD of iMs is more than  $\sqrt{2}$  as large as the observed between-person SD of iSDs, then the iMs have higher reliability than the iSDs. An overview of the ratios of between-person SDs in iMs and iSDs from published studies shows that this generally seems to be the case (see Figure 8).

Error variances of both iMs and iSDs reduce with increasing numbers of occasions, so that the corresponding reliabilities converge relatively quickly. For large numbers of occasions, differences in reliability will influence results only very little. For small numbers of occasions, however, careful attention needs to be paid to the issue of reliability before drawing any conclusions about the relative predictive validities of iMs and iSDs.

### Outlook

Research on cognitive aging is marked by a growing interest in intraindividual variability (e.g., Hultsch & MacDonald, 2004; Lövdén et al., 2007; MacDonald, Nyberg, & Bäckman, 2006; Ratcliff et al., 2006b). This interest moves the field closer to an understanding of developing individuals as dynamic systems (e.g., Nesselroade, 1991). To move even further in this direction, intraindividual variability should no longer be conceived as a supplement

<sup>2</sup> The simulation presented here was based on the assumption of normally distributed variability. For nonnormal distributions the ratio of the SEs of iMs and iSDs might change. For example, using a skewed ex-Gaussian distribution, this ratio is greater than one for all numbers of observations. For such distributions, it therefore seems that estimating iSDs is more difficult than in normally distributed cases. We thank an anonymous reviewer for pointing out that the distribution of the variables might play an important role here.



of mean performance. Rather, central tendency and dispersion should be treated as equally important characteristics of developing cognitive systems. The diffusion model implements this approach by simultaneously modeling means and variances in both accuracies and latencies. In the future, the current emphasis on means and variances may be complemented by methods and theories that retain the temporal structure of the data (see Newell, Mayer-Kress, & Liu, 2009). In either case, theoretical process models fitted to individual data will play a prominent role in this process, as they provide mechanistic explanations of individual differences and age-related changes in the relations among multiple dimensions of behavior.

## References

- Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, *6*, 259–290.
- Allaire, J. C., & Marsiske, M. (2005). Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychology and Aging*, *20*, 390–401.
- Anstey, K. J., Dear, K., Christensen, H., & Jorm, A. F. (2005). Biomarkers, health, lifestyle and demographic variables as correlates of reaction time performance in early, middle, and late adulthood. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *58A*, 5–21.
- Bassano, D., & Van Geert, P. (2007). Modeling continuity and discontinuity in utterance length: A quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science*, *10*, 588–612.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.
- Brose, A., Schmiedek, F., Lövdén, M., Molenaar, P. C. M., & Lindenberger, U. (2009). *Adult age differences in co-variation of motivation and working memory performance: Contrasting between-person and within-person findings*. Manuscript submitted for publication.
- Cattell, R. B. (1952). The three basic factor-analytic research designs: Their interrelations and derivatives. *Psychological Bulletin*, *49*, 499–520.
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997, April 10). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*, 604–608.
- Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging, Neuropsychology, and Cognition*, *12*, 187–215.
- Fiske, D. W., & Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin*, *52*, 217–251.
- Fozard, J. L., Verduyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore Longitudinal Study of Aging. *Journals of Gerontology, Series A: Psychological Sciences and Social Sciences*, *49*, P179–P189.
- Guilford, J. P. (1956). *Fundamental statistics in psychology and education*. New York, NY: McGraw-Hill.
- Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, *44*, 461–465.
- Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*, 627–634.
- Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1992). Intraindividual change in text recall of the elderly. *Brain and Language*, *42*, 248–269.
- Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, *42*, 609–629.
- Hultsch, D. F., & MacDonald, S. W. S. (2004). Intraindividual variability in performance as a theoretical window onto cognitive aging. In R. A. Dixon, L. Bäckman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 65–88). Oxford, England: Oxford University Press.
- Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, *13*, 869–881.
- Lecerf, T., Ghisletta, P., & Jouffray, C. (2004). Intraindividual variability and level of performance in four visuo-spatial working memory tasks. *Swiss Journal of Psychology*, *63*, 261–272.
- Li, S. C., Aggen, S., Nesselroade, J. R., & Baltes, P. B. (2001). Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance. *Gerontology*, *47*, 100–116.
- Li, S. C., Huxhold, O., & Schmiedek, F. (2004). Aging and attenuated processing robustness: Evidence from cognitive and sensorimotor functioning. *Gerontology*, *50*, 28–34.
- Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, *15*, 155–163.
- Li, S.-C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences*, *5*, 479–486.
- Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging*, *23*, 731–742.
- Lindenberger, U., Li, S.-C., Lövdén, M., & Schmiedek, F. (2007). The Center for Lifespan Psychology at the Max Planck Institute for Human Development: Overview of conceptual agenda and illustration of research activities. *International Journal of Psychology*, *42*, 229–242.
- Lindenberger, U., & von Oertzen, T. (2006). Variability in cognitive aging: From taxonomy to theory. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 297–314). Oxford, England: Oxford University Press.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2007). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.
- Lövdén, M., Li, S.-C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. *Neuropsychologia*, *45*, 2827–2838.
- MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria longitudinal study. *Psychology and Aging*, *18*, 510–523.
- MacDonald, S. W. S., Nyberg, L., & Bäckman, L. (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences*, *29*, 474–479.
- Martin, M., & Hofer, S. M. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology*, *50*, 7–11.
- McNemar, Q. (1962). *Psychological statistics*. New York, NY: Wiley.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201–218.
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*, 112–117.
- Nesselroade, J. R. (1984). Concepts of intraindividual variability and change: Impressions of Cattell's influence on lifespan developmental psychology. *Multivariate Behavioral Research*, *19*, 269–286.
- Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In R. M. Downs, L. S. Liben, & D. S. Palermo (Eds.), *Visions of aesthetics, the environment, & development: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale, NJ: Erlbaum.

- Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 59*, P49–P55.
- Newell, K. M., Mayer-Kress, G., & Liu, Y.-T. (2009). Aging, time scales, and sensorimotor variability. *Psychology and Aging, 24*, 809–818.
- Rabbitt, P., Osman, P., Moore, B., & Stollery, B. (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54A*, 981–1003.
- Ram, N., Rabbitt, P., Stollery, B., & Nesselroade, J. R. (2005). Cognitive performance inconsistency: Intraindividual change and variability. *Psychology and Aging, 20*, 623–633.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.
- Ratcliff, R., & Rouder, J. F. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347–356.
- Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*, 408–424.
- Ratcliff, R., Thapar, A., & McKoon, G. (2006a). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review, 13*, 626–635.
- Ratcliff, R., Thapar, A., & McKoon, G. (2006b). Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging, 21*, 353–371.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. New York, NY: Cambridge University Press.
- Salthouse, T. A. (1993). Attentional blocks are not responsible for age-related slowing. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 48*, P263–P270.
- Salthouse, T. A., & Berish, D. E. (2005). Correlates of within-person (across-occasion) variability in reaction time. *Neuropsychology, 19*, 77–87.
- Salthouse, T. A., Nesselroade, J. R., & Berish, D. E. (2006). Short-term variability and the calibration of change. *Journal of Gerontology: Psychological Sciences, 61*, 144–151.
- SAS Institute. (2006). *The GLIMMIX procedure*. Retrieved from <http://support.sas.com/rnd/app/papers/glimmix.pdf>
- Schaie, K. W. (1962). A field-theory approach to age changes in cognitive behavior. *Vita Humana, 5*, 129–141.
- Schmiedek, F. (2006, April). *The dark side of the mean*. Paper presented at the 11th Cognitive Aging Conference, Atlanta, GA.
- Schmiedek, F., Li, S.-C., & Lindenberger, U. (2009). Interference and facilitation in spatial working memory: Age-associated differences in lure effects in the *N*-back paradigm. *Psychology and Aging, 24*, 203–210.
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136*, 414–429.
- Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*, 369–385.
- Shammi, P., Bosman, E., & Stuss, D. T. (1998). Aging and variability in performance. *Aging, Neuropsychology, and Cognition, 5*, 1–13.
- Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions, 4*, 1–5.
- Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging, 21*, 545–557.
- Snijders, T. A. B., & Bosker, R. (1999). *Multilevel analysis*. Thousand Oaks, CA: Sage.
- Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 461–479.
- van der Maas, H. L., & Molenaar, P. C. M. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review, 99*, 395–417.
- Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review, 114*, 830–841.
- Wagenmakers, E.-J., Grasman, R. P. P. P., & Molenaar, P. C. M. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology, 49*, 195–204.
- Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*, 3–22.
- West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition, 49*, 402–419.
- Williams, B. R., Hultsch, D. F., Strauss, E. H., Hunter, M. A., & Tannock, R. (2005). Inconsistency in reaction time across the life span. *Neuropsychology, 19*, 88–96.

(Appendixes follow)

## Appendix A

## Penalized Radial Spline Smoothing With Mixed Models

Before being submitted to the variance heterogeneity models, observed data were fitted with the penalized radial spline smoothing method as proposed by Ruppert, Wand, and Carroll (2003) and implemented in SAS PROC GLIMMIX (SAS Institute, 2006). The radial spline smoothing method is based on the model

$$Y_{it} = \sum_{k=1}^K u_{ki}(t - \kappa_k)_+ + e_{it}, \quad (\text{A1})$$

where  $Y_{it}$  is the observed mean RT of participant  $i$  in session  $t$  and the  $(t - \kappa_k)_+$  are a set of radial basis functions ( $t - \kappa_k$  if  $t > \kappa_k$ ; 0 otherwise) with  $k$  knots  $\kappa_k$ . Knots are chosen based on a  $k$ - $d$  tree procedure, which partitions the space of all observations (number of individuals times number of sessions) until all partitions contain at most  $b$  observations. This number  $b$  is called “bucket size” and controls the number of knots. The residuals  $e_{it}$  are assumed to be distributed normally with variance  $\sigma_e^2$ . The  $u_{ki}$  are the regression weights for the  $i$ th individual

and  $k$ th knot. These weights are not estimated directly; rather their total variance (across individuals and knots) is estimated as a random variance parameter  $\sigma_u^2$  in a mixed model (see Ruppert et al., 2003, Chapter 4). This way, the amount of smoothing, which depends on the variance of the  $u_{ki}$ , is automatically chosen by the usual maximum likelihood estimation procedures of mixed models. The smoothing parameter  $\lambda$  is implicitly determined by  $\lambda^2 = \sigma_e^2/\sigma_u^2$  and thereby the same across individuals. Intercept and linear trend are included as fixed effects in the model. Fitted trends for individuals can be derived from *best linear unbiased predictors* (BLUPs) of the unobserved  $u_{ki}$ , which can be created as standard output from mixed models.

In the presented analyses, this mixed model spline smoothing approach was fitted separately to the younger and older samples, which resulted in the selection of 17 knots for both age groups. Additional analyses conducted for the combined sample resulted in virtually indistinguishable fitted functions.

The SAS code for these analyses is shown in Figure A1.

```
proc glimmix data=nback_rt noclprint maxopt=100;
  sessionNr=sessionNr/100; *rescale session variable to improve estimation;
  nb_mrt=nb_mrt/10000; *rescale dependent variable to improve estimation;
  t=sessionNr;
  by agegrp;
  class id;
  model rt = sessionNr /dist=normal solution; *define fixed effects and
  distribution of residuals;
  random t /solution *define spline smoother;
  type=rsmooth subject=id
  knotmethod=kdtree(bucket=1000 knotinfo treeinfo);
  output out= nback_rt_pred_spline pred(blup)=p_rt; *save BLUPs;
  nloptions tech=trureg; *choose optimization technique;
run;
```

Figure A1. SAS code for radial spline smoothing method as implemented in SAS PROC GLIMMIX.

## Appendix B

## Results for Variance Heterogeneity Models When Individual Trends Were Fitted With Exponential Functions

Table B1  
Results From Fitting POM and Log-Linear Variance Heterogeneity Models

Variable	POM	Log-linear
	Baseline	
Exponent	2.29	4.11
Scaling factor	0.01373	0.00029
Number of parameters	2	2
Model fit (-2LL)	-69,515	-68,054
	Different exponents for age groups	
Exponent (younger)	2.28	6.33
Exponent (older)	2.42	4.58
Scaling factor	0.01411	0.00017
Number of parameters	3	3
Model fit (-2LL)	-69,547	-69,292
	Different exponents and scaling factors for age groups	
Exponent (younger)	3.95	10.98
Exponent (older)	1.88	3.12
Scaling factor (younger)	0.10190	0.00003
Scaling factor (older)	0.00878	0.00045
Number of parameters	4	4
Model fit (-2LL)	-71,255	-71,097

*Note.* Separate fits for individuals are not reported because parameter estimation was not successful for too many individuals. POM = power of means.

As an alternative to the spline smoothing method, individual trends were also fitted separately for each individual using the three-parameter exponential function

$$Y_t = a + ge^{-r(t-1)}, \quad (\text{B1})$$

with  $Y_t$  being the observed RT mean at session  $t$ ,  $a$  the asymptote parameter,  $g$  the gain parameter, and  $r$  the rate parameter of

the exponential function ( $e$ ). Individual functions were fitted using a least-square procedure as implemented in SAS PROC NLIN. Predicted values from these fitted functions and residual variability around them were used to define mean performance and intraindividual variability in the variance heterogeneity models shown in Table B1. The SAS code for these analyses is shown in Figure B1.

```
proc nlin data=nback_rt;
  by id;
  parms asymp=200 gain=500 rate=.3;          *starting values;
  model rt = asymp + gain*exp(-rate*(sessionnr-1)); *exponential model;
  output out = nb_rt_pred_exp predicted=epr_nb_rt; *save predicted values;
run;
```

Figure B1. SAS code for least-square procedure as implemented in SAS PROC NLIN.

Received January 7, 2009  
Revision received June 16, 2009  
Accepted September 25, 2009 ■