

# Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study

Martin Lövdén\*, Shu-Chen Li, Yee Lee Shing, Ulman Lindenberger

*Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, D-14195 Berlin, Germany*

Received 17 October 2006; received in revised form 6 April 2007; accepted 7 May 2007

Available online 17 May 2007

## Abstract

Neurocomputational modeling and empirical evidence suggest that losses in neuronal signaling fidelity cause senescent changes in behavior. We applied structural equation modeling to five-occasion 13-year longitudinal data from the Berlin Aging Study ( $n = 447$ ; age range at  $t_1 = 70$ – $102$  years) to test whether trial-to-trial reaction time variability in perceptual speed (identical pictures) antecedes and signals longitudinal decline in levels of performance on perceptual speed (digit letter and identical pictures) and ideational fluency (category fluency). Higher trial-to-trial variability preceded and predicted greater cognitive decline in perceptual speed and ideational fluency. We conclude that trial-to-trial variability signals impending decline in cognitive performance, and that theories of neurocognitive aging need to postulate developmental cascades between senescent changes in variability and central tendency.

© 2007 Elsevier Ltd. All rights reserved.

**Keywords:** Within-person variability; Inconsistency; Neuronal noise; Longitudinal change; Neurocognitive aging; Cognitive control

The dynamics of change and its underlying mechanisms are fundamental to the understanding of behavior, development, and aging (Baltes, Reese, & Nesselroade, 1977; Hultsch & MacDonald, 2004; Li, Huxhold, & Schmiedek, 2004a; Lindenberger & Oertzen, 2006; MacDonald, Nyberg, & Bäckman, 2006; Nesselroade & Boker, 1994). Nesselroade (1991) (see also Li, Huxhold et al., 2004; Li, Lindenberger et al., 2004) differentiated between two types of within-person change: (i) intraindividual change denotes relatively permanent alterations that evolve slowly over relatively long-term time scales (e.g., months and years) and (ii) intraindividual variability (or, inconsistency) denotes transient and rapid fluctuations that occur over shorter time scales (e.g., moments and days). This study examines dynamic relations between intraindividual changes at these long-term and short-term time scales. First, we model the longitudinal change trajectories of both intraindividual cogni-

tive variability and of mean levels of performance in old and very old age. Second, we test whether short-term intraindividual variability antecedes and signals impending long-term cognitive decline.

Accumulating evidence suggests that short-term variability in cognitive performance displays relatively stable between-person differences over time (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000; Rabbitt, Osman, Moore, & Stollery, 2001; Ram, Rabbitt, Stollery, & Nesselroade, 2005), across cognitive tasks (Fuentes, Hunter, Strauss, & Hultsch, 2001; Hultsch et al., 2000), and across domains of functioning (Li, Aggen, Nesselroade, & Baltes, 2001). Behavioral or neuronal indicators point to higher variability in persons with dementia (e.g., Hultsch et al., 2000), in schizophrenic patients (e.g., Winterer & Weinberger, 2004), and after brain injury (Stuss, Pogue, Buckle, & Bondar, 1994; West, Murphy, Armilio, Craik, & Stuss, 2002). Importantly, adult age-related increases in cognitive variability have been observed cross-sectionally (Li, Huxhold et al., 2004; Li, Lindenberger et al., 2004; Spieler, Balota, & Faust, 1996; Williams, Hultsch, Strauss, Hunter, & Tannock, 2005; but see Robertson, Myerson, & Hale, 2006; Shammi, Bosman, & Stuss, 1998) and longitudinally (Deary &

\* Corresponding author.

*E-mail addresses:* loevden@mpib-berlin.mpg.de (M. Lövdén), Shuchen@mpib-berlin.mpg.de (S.-C. Li), YShing@mpib-berlin.mpg.de (Y.L. Shing), Lindenberger@mpib-berlin.mpg.de (U. Lindenberger).

Der, 2005; Fozard, Verduyssen, Reynolds, Hancock, & Quilter, 1994; MacDonald, Hultsch, & Dixon, 2003). Moreover, individuals with greater variability perform less well on a wide variety of cognitive tasks (Li, Aggen et al., 2001; Li, Lindenberger, & Sikstrom, 2001; Rabbitt et al., 2001; Rabbitt, 2000; Salthouse, 1993), including measures of fluid intelligence and working memory (Li, Aggen et al., 2001; Li, Lindenberger et al., 2001; Li, Huxhold et al., 2004; Li, Lindenberger et al., 2004; Schmiedek, Oberauer, Wilhelm, Süß, & Wittman, in press; but see Allaire & Marsiske, 2005). Moreover, after controlling for central tendency, cognitive variability still accounts for substantial amounts of variance in dependent variables such as age (e.g., Hultsch, MacDonald, & Dixon, 2002; MacDonald et al., 2003; Williams et al., 2005; but see Deary & Der, 2005; Salthouse, 1993), cognitive functioning (Li, Aggen et al., 2001; Li, Lindenberger et al., 2001; Rabbitt, 2000; Ram et al., 2005; but see Salthouse & Berish, 2005), and disease (Strauss, MacDonald, Hunter, Moll, & Hultsch, 2002).

In an important study, MacDonald et al. (2003) reported age-related changes (over 6 years) in within-person reaction time variability for a group of individuals aged 75–89 years. Sizeable correlations between changes in variability and changes in performance levels on tasks assessing working memory, perceptual speed, episodic memory, vocabulary, and reasoning were observed. In sum, cognitive intraindividual variability is a relatively stable trait, vulnerable to neurological disease and aging, related to cognitive performance, and associated with decline in cognitive functioning in old age (e.g., MacDonald et al., 2006).

Cognitive moment-to-moment variability has been linked to the integrity of executive functions, working memory, and cognitive control (e.g., Stuss, Murphy, Binns, & Alexander, 2003; West et al., 2002). For example, findings indicate that age-related increases in reaction time variability reside, in part, in the slow end of the reaction time distribution (Hultsch et al., 2002; Schmiedek et al., in press; West et al., 2002; Williams et al., 2005). These findings have been taken to suggest that specific processes related to cognitive control, such as attentional lapses (Bunce, Warr, & Cochrane, 1993) and robustness of maintenance of task context (e.g., Braver et al., 2001), are involved in age differences in cognitive variability (Hultsch et al., 2002; West et al., 2002; Williams et al., 2005; but see Schmiedek et al., in press). This line of reasoning is supported by pronounced age-related structural (e.g., Sullivan & Pfefferbaum, 2006) and functional alterations (Buckner, 2004) in the integrity of prefrontal cortex, an area of the brain that supports cognitive control. Direct support for the link between prefrontal cortex and cognitive variability comes from functional imaging work showing an positive association between between-person differences in reaction time variability and brain activity in bilateral middle frontal areas during a response inhibition task (Bellgrove, Hester, & Garavan, 2004). Also, psychiatric pathologies that affect cognitive control, such as schizophrenia, are associated with increased neuronal noise (e.g., Winterer & Weinberger, 2004). In sum, increased moment-to-moment variability is currently most strongly related to the prefrontal cortex and cognitive control processes (e.g., MacDonald et al., 2006).

A related line of theoretical work formally connects cognitive variability to increased neural noise originating from age-related reduction in the efficacy of catecholaminergic, notably dopaminergic, neuromodulation (Li, Brehmer, Shing, Werkle-Bergner, & Lindenberger, 2006; Li & Lindenberger, 1999; Li, Lindenberger et al., 2001). The model by Li et al. provided an integrated account of deficient neuromodulation and several cognitive aging phenomena. First, the model recounts and provides a formalism for earlier hypotheses that cognitive decline is driven by increased information loss owing to increasing neural noise (Hendrickson, 1982; Welford, 1965), and specifically relates increases in neural noise to reductions in dopamine transmitter content and binding functionality (Bäckman, Nyberg, Lindenberger, Li, & Farde, 2006). Second, the model accounts for findings that point to age-related decrease in the distinctiveness of cortical representations. For example, older adults tend to encode information less distinctively ( Craik, 1983) and more variably (Mäntylä & Bäckman, 1990) than young adults. Moreover, correlations among performances on different cognitive tests increase in groups of older adults (Baltes & Lindenberger, 1997; Li, Huxhold et al., 2004; Li, Lindenberger et al., 2004). Recently, neuroimaging work has lent further and more direct empirical support for this notion (Park et al., 2004). In turn, less distinctive neural representations may produce more behavioral variability (see also Thaler, 2002). Li et al. (Li & Lindenberger, 1999; Li, Lindenberger, & Frensch, 2000; Li, Aggen et al., 2001; Li, Lindenberger et al., 2001; Li, Brehmer et al., 2006; Li, Oertzen, & Lindenberger, 2006; Li, Naveh-Benjamin, & Lindenberger, 2005; Li & Sikstrom, 2002; Li, Oertzen et al., 2006) have successfully modeled these, and other, cognitive aging phenomena by adjusting the gain parameter of the activation function of neural networks, simulating a reduced signal-to-noise ratio of neural information processing in the aging brain that is attributable to deficient neuromodulation.

The reviewed strands of neurocomputational modeling work and empirical evidence have stimulated the speculation that decreasing processing fidelity, as captured by increasing within-person reaction time variability, antecedes and signals cognitive decline in late adulthood and old age (e.g., Eizenman, Nesselroade, Featherman, & Rowe, 1997; Li & Lindenberger, 1999; Lindenberger, Li, & Bäckman, 2006; cf. Van der Maas & Molenaar, 1992; Siegler, 1994). As Hultsch and MacDonald (2004) pointed out, the identification of increasing variability as an early warning signal or, figuratively speaking, “canary in the coal mine,” would carry significant clinical implications.

Accordingly, the main goal of this study was to examine whether higher levels of intraindividual trial-to-trial reaction time variability signal subsequent negative longitudinal changes in cognitive performance (see Fig. 1). To this end, we used a type of structural equation model that is particularly amenable for investigating temporal lead–lag dynamics between antecedents and outcomes of ontogenetic change. Specifically, we made use of the dual change score model (DSCM) originally proposed by McArdle and Hamagami (2001). The DSCM has been fruitfully applied to structurally analogous developmental hypotheses (Ferrer & McArdle, 2004; Ghisletta & de Ribaupierre, 2005; Ghisletta & Lindenberger, 2003; Ghisletta & Lindenberger,

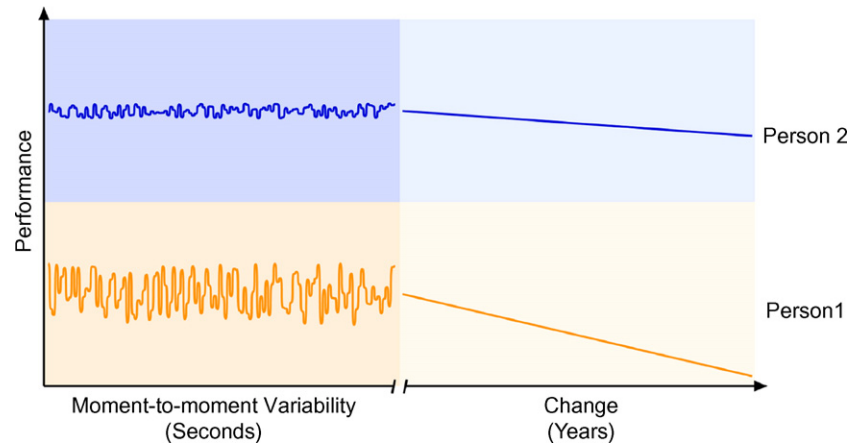


Fig. 1. Illustration of guiding prediction. Aging individuals with greater moment-to-moment process fluctuations at a given point in time are expected to show greater subsequent longitudinal decline in mean levels of functioning than individuals who fluctuate less. Adapted from Lindenberger et al. (2006).

2005; Ghisletta, Bickel, & Lövdén, 2006; Lövdén, Ghisletta, & Lindenberger, 2005; McArdle et al., 2004). The DCSM represents a dynamic extension of standard latent growth curve models (LGCM; Meredith & Tisak, 1990) or longitudinal multi-level models (Bryk & Raudenbush, 1987; for a comparison, see Lindenberger & Ghisletta, 2004). Like the LGCM, the DCSM estimates means and variances for the intercept (i.e., initial level) and slope (i.e., change) of a longitudinal time series at the latent level that are unbiased by measurement error. Residual terms (e.g., errors) are estimated simultaneously and separately. Unlike the LGCM, the DCSM permits estimation of time-lagged (or dynamic) relations between a variable's level at one time point and its subsequent change. In its multivariate form, the model allows for corresponding time-lagged associations between different variables. In contrast to typical cross-lagged correlations (Rogosa, 1980), the multivariate extension of the DCSM permits adjustment for unequal reliabilities of the variables, separation of within-variable and across-variable lead-lag effects, and direct and simultaneous modeling of the variables' changes.

With the multivariate DCSM, we can empirically examine, in a precise manner, whether cognitive variability precedes and predicts decline cognitive performance, and, importantly, we can pit this hypothesis against the alternative that cognitive performance precedes and predicts changes in cognitive variability. The data used in this study come from the Berlin Aging Study (BASE; Baltes & Mayer, 1999), and refer to a total of 447 individuals aged 70–102 years who were observed for up to five occasions over up to 13 years. Measures include trial-to-trial variability in reaction time performance and typical indicators of cognitive performance in old age (perceptual speed and ideational fluency).

Our analyses of this data set addressed two questions. First, we examined whether changes in intraindividual (trial-to-trial) variability are associated with performance changes in mean levels of perceptual speed and ideational fluency. Here, we attempt to replicate the findings from MacDonald et al. (2003). Second, with bivariate DCSMs, we investigate lead-lag relations between within-person reaction time variability and the measures of performance level. Thus, we extend the work of

MacDonald et al. by explicitly comparing the temporal order of senescent changes in variability and central tendency.

## 1. Method

So far, the interdisciplinary multi-session longitudinal BASE study has collected longitudinal data over 13 years. The baseline assessment (t1) took place in 1990–1993. The second assessment (t2) was conducted 1.95 years ( $SD=0.71$ ), t3 3.76 years ( $SD=0.66$ ), t4 5.53 years ( $SD=0.79$ ), t5 8.94 years ( $SD=0.84$ ), and t6 13.00 years ( $SD=0.87$ ) after t1, respectively. Only a reduced assessment protocol was completed at t2, therefore this assessment is not considered here. Detailed descriptions of the design, sample, procedures, and measures have been published elsewhere (Baltes & Mayer, 1999; Lövdén, Ghisletta, & Lindenberger, 2004). A brief description is provided below.

### 1.1. Participants

To obtain the t1 sample, 1,908 individuals, drawn from the Berlin city registry, were approached. Initial response rate was 78% ( $n=1,491$ ) and, out of those, 516 individuals completed a 14-session assessment protocol (age range = 70–103 years;  $M=84.9$  years;  $SD=8.7$ ). Selectivity analyses comparing this t1 sample with the total parent sample showed that the t1 sample was positively selected on variables covering demographic, sensory, life history, and intellectual domains (Lindenberger et al., 1999). However, with the exception of dementia prevalence, effect sizes were small (i.e., below  $0.5SD$ ). Furthermore, variances and covariances were only marginally influenced by selectivity. Thus, the t1 sample was reasonably representative of the target population.

Out of these 516 individuals, 69 had severe vision problems preventing them from completing the normal cognitive measurement battery and consequently lacked data on trial-to-trial variability. Therefore, we included the remaining 447 participants only (age range = 70–102 years;  $M=84.1$  years;  $SD=8.5$ ). Expressed over the  $SD$  of the total t1 sample ( $n=516$ ), this remaining sample was not positively selected on other important dimensions of this study: chronological age ( $0.1SD$ ), ideational fluency ( $0.1SD$ ), and vocabulary ( $0.0SD$ ).

Longitudinal mean selectivity for the included measures was analyzed and expressed in an effect size metric (see Lindenberger, Singer, & Baltes, 2002), indicating the magnitude to which individuals that survived and participated in t6 ( $n=47$ ) differed from the t1 sample ( $n=447$ ). Total selectivity at t6 for measures assessed at t1 was  $0.9SD$  for chronological age,  $0.8SD$  for perceptual speed,  $0.9SD$  for ideational fluency,  $0.5SD$  for vocabulary, and  $0.3SD$  for trial-to-trial variability. To the extent that selectivity is captured by the variables included in the analyses, the estimation algorithm used in this study accounts for selective attrition when estimating the population parameters (see Section 1.3).

Based on age-specific cutoffs on the short mini-mental state examination (SMMSE; Klein et al., 1985), 118 of the participating individuals at t1 (total

$n=447$ ), 51 at t3 ( $n=191$ ), 30 at t4 ( $n=124$ ), 16 at t5 ( $n=78$ ), and six at t6 ( $n=46$ ) were identified as suspected dementia cases. The cutoff was 12 points on the SMMSE for 70–84 year-olds and 11 points for older individuals. Based on independent clinical diagnoses of dementia that were issued at t1 and t3 (using standard assessment procedures; Helmchen et al., 1999) these cutoffs resulted in acceptable specificity (>72% for all subgroups, including younger and older men and women) and sensitivity (>63% for all subgroups; see Gerstorff, Herlitz, & Smith, 2006 for details). The way suspected dementia is handled in the analyses are described in Section 1.3.

1.2. Measures

Medical personnel or trained research assistants carried out all testing in individual sessions at the participants' place of residence. Data collection for the cognitive measures was supported by a Macintosh SE/30 equipped with a touch-sensitive screen. The sessions lasted for an average of 90 min.

The complete longitudinal cognitive battery of BASE consists of two perceptual speed tasks (identical pictures and digit letter), two episodic memory tasks (memory for text and paired associates), two fluency tasks (categories and word beginnings), and two verbal knowledge tasks (vocabulary and spot-a-word; see Lindenberger, Mayr, & Kliegl, 1993; Lövdén et al., 2004, for details). For this study we initially considered the performance measures from the full cognitive battery and trial-to-trial reaction time variability in the identical pictures task. All these measures were initially analyzed with univariate linear LGCMs using chronological age as a covariate (see Section 1.3). Because the main multivariate analyses estimate covariances between changes in different variables, we selected, for this study, the measures that evinced significant interindividual differences in change (i.e., slope) at the .05 level. This is due to the fact that variance in change is a prerequisite for estimating covariances among changes. Based on this criterion, digit letter, categories, and trial-to-trial variability were included.

1.2.1. Category fluency

For this ideational fluency task, participants named as many animals as possible within 90 s. Responses were tape-recorded. The performance measure refers to the number of correct responses (i.e., no repetitions, right category, no morphological variations).

1.2.2. Digit letter

This measure of perceptual speed corresponds closely to the well-known digit symbol substitution of the WAIS. However, participants verbally named the letters instead of writing symbols. The reason for this change was to minimize motor demands. A template with digit–letter pairings was visible during the entire testing period. Each one of a total of 21 sheets contained six digits with question marks beneath. Participants moved from left to right, naming the letters corresponding to the digits. The dependent measure was the number of correct responses after 3 min.

1.2.3. Trial-to-trial variability (identical pictures)

Our measure of trial-to-trial variability in cognitive performance is derived from a test of perceptual speed: the identical pictures. This measure was selected because it was the only one for which reaction times were recorded for each of the trials separately. A selection of items from the corresponding ETS test (Ekstrom, French, Harman, & Derman, 1976) was computerized and used for this test. For each item, a target figure in the upper part and five response alternatives in the lower part of the screen were presented. Participants touched, as fast as possible, the one alternative figure that corresponded to the target figure. Presentation of the items terminated after 80 s. Thus, there were individual differences (within and across assessments) in the number of items each individual responded to.

Because the research community has not established a standard for calculating intraindividual trial-to-trial variability, we computed several variability measures: (a) the intraindividual SD based on all available (correct) responses for each individual and each assessment; (b) the intraindividual coefficient of variation (CoV;  $SD/M$ ) based on all available (correct) responses for each individual and each assessment; (c) the intraindividual SD based on the first 10 trials only (incorrect responses treated as missing); (d) the CoV based on the first 10 trials only (incorrect responses treated as missing); (e) the intraindividual SD after the average effects of trial and chronological age were statistically partialled out from 10 trials of raw data (after imputation of missing values; cf. MacDonald et al., 2003); and (f) the residual intraindividual SD based on all available raw data after statistically partialling out the mean. Before each of these computations, outliers in the raw data (<2%) were deleted. A comparison between the mean RT for the first 10 trials (3,468 ms) and for the last 10 trials (3,540 ms) suggested that no improvement occurred across trials. Thus, we found it unnecessary to detrend the data for individual learning trends. The number of items

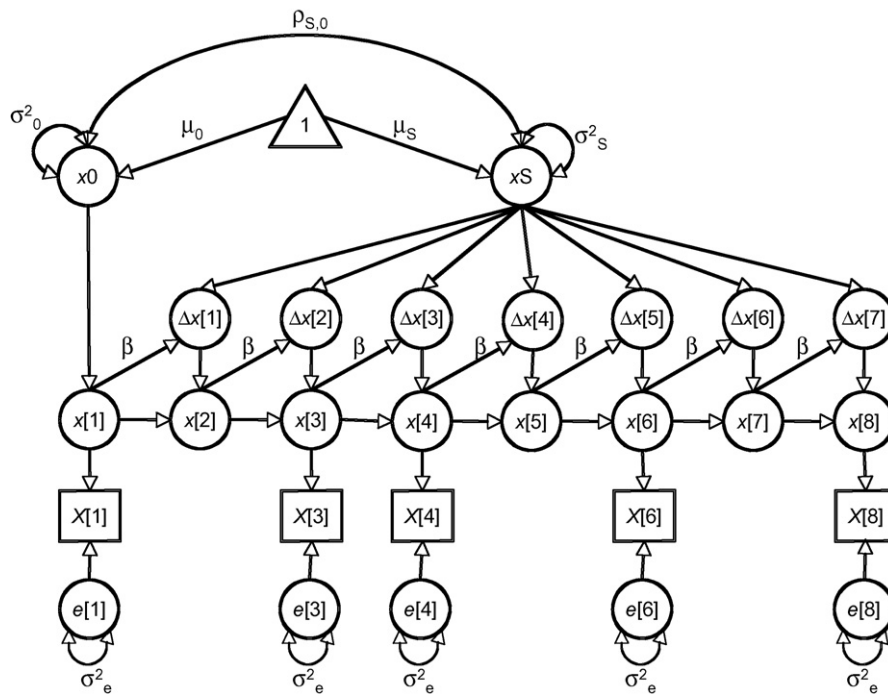


Fig. 2. Graphical representation of a univariate dual change score model (McArdle & Hamagami, 2001) as implemented here. Observed variables are depicted as squares, latent variables as circles, regressions as one-headed arrows, and variances and covariances as two-headed arrows. Means and intercepts are indicated by the triangle. Unlabeled paths are fixed to 1.



was too few to allow for modeling the reaction time distribution with formal models.

All the intraindividual trial-to-trial variability measures provided converging main results. Therefore, we report only the analyses of the intraindividual CoV based on all available trials. The statistical models including this measure produced the most stable and robust solutions (i.e., no convergence problems and no improper estimates in any of the models applied). The 4-year (t1–t3) test retest correlation for this measure was 0.34, indicating marginally decent lower boundary of reliability. For simplicity and because all measures produced comparable results, we refer to this measure as trial-to-trial variability.

1.3. Statistical procedures

The data were analyzed with structural equation models, and specifically with a multivariate LGCM (Meredith & Tisak, 1990) and bivariate DCSMs (McArdle & Hamagami, 2001). The LGCM is, generally speaking, nested under the DCSM. Therefore, we take the LGCM as starting point for describing the DCSM (see also Ferrer & McArdle, 2004; Ghisletta & de Ribaupierre, 2005; Ghisletta & Lindenberger, 2003; Ghisletta & Lindenberger, 2005; McArdle & Hamagami, 2001; McArdle et al., 2004).

Fig. 2 depicts a graphical representation of a univariate DCSM. We assume latent values of  $x[t]$ , where  $t$  is time of assessment, with 2 years between each (i.e.,  $x[1], \dots, x[8]$ ). The observed variables  $X[t]$  were measured five times (i.e., t1, t3, t4, t5, t6), with approximately 4 years between t1 and t3, 2 years between t3 and t4, 4 years between t4 and t5, and 4 years between t5 and t6 (see Table 1). Thus, translating the actual measurement intervals into the equal-intervals specified in the models, the observed variables were measured at  $X[1], X[3], X[4], X[6]$ , and  $X[8]$ . This equal-interval approach allows invariant scaling over the time series (see also Gerstorff, Lövdén, Rocke, Smith, & Lindenberger, 2007; Ghisletta & Lindenberger, 2003; Ghisletta & Lindenberger, 2005; Lövdén et al., 2005). The variables  $x[t]$  are the standardized sums of the latent score at  $x[t - 1]$  plus the difference score  $\Delta x[t]$ , so that  $\Delta x[t]$  translates into the error-free difference score between  $x[t - 1]$  and  $x[t]$  (McArdle & Nesselroade, 1994). The error term  $\sigma_e^2$  is assumed to be normally distributed with a mean of zero and to neither change

over time nor correlate with itself. The latent intercept  $x_0$  (i.e., the latent scores at t1) and linear slope  $xS$  (i.e., the linear 2-year change scores) are anticipated to account for the time series information. The means of the intercept and slope ( $\mu_0$  and  $\mu_S$ ), their variances (i.e.,  $\sigma_0^2$  and  $\sigma_S^2$ ), and their covariance  $\rho_{S,0}$  are estimated. These six parameters ( $\mu_0, \mu_S, \sigma_0^2, \sigma_S^2, \rho_{S,0}, \sigma_e^2$ ) are estimated in a linear LGCM. The DCSM additionally allows for estimation of the dynamic parameter  $\beta$ , denoting the effects of a score at  $x[t - 1]$  on the impending change between  $x[t - 1]$  and  $x[t]$ . The  $\beta$ 's are usually assumed time-invariant. With  $\beta$  fixed to zero, the DCSM is a linear LGCM.

Multivariate extensions of the LGCM and the DCSM are possible. A bivariate LGCM estimates two times the six parameters of each time series plus four covariances among the intercepts and slopes. The bivariate DCSM (BDCSM) estimates four additional, dynamic, parameters: the two  $\beta$ 's of each time series plus two corresponding inter-variable parameters, called  $\gamma$ 's, denoting the effect that a variable at  $t - 1$  exerts on the coming change of the other variable (i.e., between  $t - 1$  and  $t$ ). Fixing the  $\beta$ 's and  $\gamma$ 's to zero results in bivariate LGCM. Alternative hypotheses can be postulated by restricting the full BDCSM, which estimates both inter-variable  $\gamma$  parameters (model full coupling). Consider, for example, a BDCSM including trial-to-trial variability and categories performance. Fixing the inter-variable  $\gamma$  parameter from level of variability to changes in categories performance to zero would postulate a model in which categories performance predicts changes in variability, but not the other way round. Significant reductions in fit for this model as compared with the full coupling model means that variability significantly predicts changes in categories performance. Conversely, if fixing the  $\gamma$  parameter from level of categories performance to variability to zero results in reductions in fit, then categories performance reliably precedes and predicts changes in variability.

Time-invariant covariates are implemented in the LGCM and the DCSM as influencing the intercept  $x_0$  and the slope  $xS$  (and, via the slope, the  $\Delta x[t]$ ). In all analyses, we include two time-invariant covariates, chronological age and time-to-death. In addition, we include suspected dementia at each assessment as influencing the corresponding observed cognitive variable. For example, suspected dementia at t1 influenced the observed variable  $X[1]$ , suspected dementia at t3 influenced the observed variable  $X[3]$ , and so fourth. The suspected demen-

Table 1  
Descriptive statistics for measures included in the models

Measure	Time	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Trial-to-trial variability (CoV)	t1	443	50.0	10.0	0.6	1.0
	t3	179	51.1	9.3	0.8	0.8
	t4	118	50.0	8.8	0.7	0.5
	t5	70	49.3	8.5	0.5	0.9
	t6	40	49.3	9.2	1.0	2.0
Categories	t1	447	50.0	10.0	0.4	0.0
	t3	190	52.5	10.3	0.0	0.0
	t4	124	54.2	11.0	0.2	-0.1
	t5	77	53.9	10.9	0.2	-0.6
	t6	46	52.0	11.8	-0.2	-0.3
Digit letter	t1	440	50.0	10.0	-0.1	-0.6
	t3	172	52.8	9.5	-0.4	-0.2
	t4	117	53.4	9.5	-0.6	0.1
	t5	67	54.6	9.3	-0.9	1.0
	t6	41	50.9	8.4	-0.5	0.1
Suspected dementia	t1	447	-4.7	8.8	1.1	-0.8
	t3	191	-4.7	8.9	1.1	-0.9
	t4	124	-5.1	8.6	1.2	-0.5
	t5	78	-5.9	8.1	1.5	0.2
	t6	46	-7.4	6.8	2.3	3.3
Time-to-death	t1	342	0.0	10.0	0.7	-0.1
Chronological age	t1	447	0.0	10.0	0.1	-1.1

Note: CoV = coefficient of variation. The second time of assessment (t2) was conducted on average 1.95, t3, 3.76, t4, 5.53, t5, 8.94, and t6, 13.00 years after t1, respectively. Suspected dementia is coded as 10 and no suspected dementia as -10. All other variables are standardized to the *T*-metric (*M* = 50; *SD* = 10) with the t1 assessment providing the reference values. Time-to-death and chronological age are additionally centered.

tia variables were modeled unrelated to all other variables. In sum, the main parameters of interest (the covariances,  $\beta$ 's, and  $\gamma$ 's) are statistically controlled for the effects of chronological age, time-to-death, and suspected dementia.

We utilized AMOS 5.0 (Arbuckle & Wothke, 1999) and full information maximum likelihood (FIML; Arbuckle, 1996). FIML can be used with incomplete data under the missing-at-random assumption (Schafer & Graham, 2002). The procedure does not impute values but uses available information for estimating parameters involving variables with missing values. According to the missing-at-random assumption, the likelihood that a score on  $X$  is missing may depend on other variables in the model but not on  $X$  itself. Note that a relationship between missingness and  $X$  emanating from a common association to other variables in the model is permissible, but no residual relationship between missingness and  $X$  is allowed. Thus, FIML provide accurate estimates when the assumption is met by including satisfactory predictors of missingness. Here, the procedure is enhanced by including powerful predictors of longitudinal dropout (i.e., chronological age, time-to-death, suspected dementia, and cognitive performance; Lindenberger et al., 2002). FIML, and related approaches, produce less biased and more precise population estimates than other common procedures to missing values (e.g., listwise deletion, regression imputation; e.g., Schafer & Graham, 2002).

With FIML comes a log-likelihood fitting function ( $-2LL$ ) for the applied model and not a  $\chi^2$  value.<sup>1</sup> The difference between the  $-2LL$ s of nested models is  $\chi^2$  distributed with degrees of freedom equal to the difference in the number of parameters. The corresponding  $\chi^2$ -test was used to compare nested models. The alpha-level for all statistical decisions was 0.05.

## 2. Results

Table 1 presents descriptive statistics. All variables, with the exception of suspected dementia, were standardized (linearly transformed) to a  $T$ -score metric ( $M = 50$ ;  $SD = 10$ ), with the  $t1$ -assessment providing the reference distribution. Time-to-death and chronological age were additionally centered to a mean of zero. All variables displayed acceptable distributions (i.e., skewness and kurtosis).

### 2.1. Are changes in trial-to-trial variability associated with changes in cognitive performance levels?

We estimated a trivariate linear LGCM,<sup>2</sup> including trial-to-trial variability (CoV), digit letter, and categories. This model extracts the means and variances for the intercepts (i.e.,  $t1$ -scores) and the linear slopes (i.e., 2-year change scores) for the three variables. Most importantly, the covariances among the intercepts and slopes of the three variables are estimated. An

<sup>1</sup> With FIML, a  $\chi^2$  statistic and derived indices could be calculated from the difference between the log-likelihood functions of the saturated (unrestricted) model and the restricted model (e.g., the applied DCSM). However, in the saturated models the number of parameters is higher than the sample size and therefore not enough information was available in the data to fit the saturated models. Thus, we report only the value of the function of log-likelihood. The lack of the conventional fit indices for the models is not problematic because the major focus is on the parameter estimates and differences between alternative models. In other words, the objectives are to extract parameter estimates of interest from the time series and to compare these estimates across alternative models (see Section 2).

<sup>2</sup> Preliminary analyses also examined the orthogonal quadratic component of the time series. Because the variance (i.e., the interindividual differences) in change was not significant for any of the four variables we could not address the associations among quadratic changes, and thus the linear LGCM adequately addresses our research question.

Table 2

Trivariate latent growth curve model: estimates (S.E) of means and variances for trial-to-trial variability (CoV), categories, and digit letter, after statistically controlling for the influence of time-to-death, chronological age, and suspected dementia

Parameter	Variability (CoV)	Categories	Digit letter
Intercept mean ( $\mu_0$ )	50.9* (0.5)	48.8* (0.4)	48.5* (0.4)
Intercept variance ( $\sigma_0^2$ )	34.9* (6.7)	45.7* (4.9)	46.4* (4.2)
Slope mean ( $\mu_S$ )	1.2* (0.2)	-0.9* (0.2)	-1.4* (0.2)
Slope variance ( $\sigma_S^2$ )	1.4* (0.6)	1.6* (0.4)	1.5* (0.3)
Error variance ( $\sigma_e^2$ )	56.0* (4.4)	25.8* (2.1)	14.5* (1.2)

Note. CoV = coefficient of variation.

\*  $p < 0.05$ .

inspection of the standardized covariances (i.e., the correlations) among the slopes addresses the question whether changes in trial-to-trial variability are associated with changes in cognitive performance (digit letter and categories). Because the error variances are simultaneously and separately estimated, the estimates of substantive interest are not biased by error. As outlined in Section 1.3, we included time-to-death at  $t1$ , chronological age at  $t1$ , and suspected dementia as covariates. The  $-2LL$  (203,  $N = 447$ ) for this model, which estimated 72 parameters, was 22298.8.

Estimates of means and variances are reported in Table 2. All slope means were significant, demonstrating average increases in variability and declines in cognitive performance. The mean linear 2-year change is greatest for digit letter ( $-1.4$ ), followed by trial-to-trial variability (1.2), and categories ( $-0.9$ ). All slope variances were significant, indicating reliable between-person differences in change.

Correlations among intercepts and slopes are reported in Table 3. Changes among the cognitive variables are highly correlated. Most importantly, changes in trial-to-trial variability, as indicated by the CoV, are significantly associated with changes in categories ( $-0.82$ ) and digit letter ( $-0.68$ ), indicating that increments in within-person trial-to-trial variability are associated with decrements in cognitive performance.

Concerning relations to other variables, the decision to include the covariates was supported. Older age significantly predicted higher initial variability and lower initial cognitive

Table 3

Are changes in trial-to-trial variability (CoV) associated with changes in mean cognitive performance?

Parameter	1	2	3	4	5	6
1. IC variability	–					
2. IC categories	-0.25 <sup>a</sup>	–				
3. IC digit letter	-0.37 <sup>a</sup>	0.63 <sup>a</sup>	–			
4. SL variability	-0.55 <sup>a</sup>	0.16	0.25	–		
5. SL categories	0.07	0.02	-0.11	-0.82 <sup>a</sup>	–	
6. SL digit letter	-0.24	0.11	0.09	-0.68 <sup>a</sup>	0.98 <sup>a</sup>	–

Correlations among intercepts (IC) and slopes (SL) of trial-to-trial variability, categories, and digit letter, after statistically controlling for the influence of time-to-death, chronological age, and suspected dementia. Note: CoV = coefficient of variation; IC = intercept (i.e., scores at  $t1$ ). SL = slope (i.e., 2-year linear change scores). Variability = trial-to-trial variability. Significance testing refers to the underlying covariances.

<sup>a</sup>  $p = 0.05$ .

performance. Moreover, older age reliably predicted greater increase in variability and stronger decreases in digit letter performance. Individuals in closer proximity to death had lower initial cognitive performance. Suspected dementia predicted lower digit letter performance at t1 to t6, lower categories performance at t1 to t5, and higher trial-to-trial variability at t1 and t6.

In summary, increases in trial-to-trial variability were associated with decreases in cognitive performance, thereby replicating the findings from MacDonald et al. (2003).

## 2.2. Is high trial-to-trial variability preceding and predicting decline in cognitive performance levels?

We initially estimated two BDCSMs (full coupling model): one model included trial-to-trial variability (CoV) and categories and one included CoV and digit letter. Time-to-death, chronological age, and suspected dementia were covariates. Each of these models estimates four parameters that have informative value over and above the estimates from the multivariate LGCM. The estimates of the  $\beta$  parameters, one for each variable, reflect the dynamic influence of level of one variable on the subsequent change of the same variable. The estimates of the two  $\gamma$  parameters reflects the dynamic influence of level of  $X$  on the subsequent change of  $Y$  and the influence of level of the  $Y$  on the subsequent change of  $X$ , respectively. Thus, the estimates of the  $\gamma$  parameters speak to the issue whether high trial-to-trial variability is preceding and predicting decline in cognitive performance levels, or vice versa. The estimates of the  $\beta$  and  $\gamma$  parameters for all BDCSMs are displayed in Table 4.

The  $-2LL$  (117,  $N=447$ ) for the model including trial-to-trial variability (CoV) and categories, which estimated 53 parameters, was 18,472.2. The influence of variability on subsequent change of categories performance ( $\gamma_{\text{variability} \rightarrow \text{categories}}$ ) was negative ( $-0.84$ ). The influence of categories on subsequent change of variability ( $\gamma_{\text{categories} \rightarrow \text{variability}}$ ) was also negative but weaker ( $-0.11$ ). Fixing the  $\gamma_{\text{variability} \rightarrow \text{categories}}$  parameter to zero resulted in a significant loss of fit,  $\chi^2(1, N=447)=11.3$ . In contrast, fixing the  $\gamma_{\text{categories} \rightarrow \text{variability}}$  parameter to zero resulted in negligible loss of fit,  $\chi^2(1, N=447)=0.1$ . Thus, the results indicate that higher trial-to-trial variability reliably precedes and predicts greater 2-year negative decline in categories performance, whereas categories performance has negligible influence on change in variability.

The  $-2LL$  (117,  $N=447$ ) for the model including trial-to-trial variability (CoV) and digit letter, which estimated 53 parameters, was 17971.6. The estimate of the inter-variable  $\gamma_{\text{variability} \rightarrow \text{digit letter}}$  parameter was negative ( $-1.30$ ) and fixing this parameters to zero resulted in a significant loss of fit,  $\chi^2(1, N=447)=25.3$ . The estimate of the inter-variable  $\gamma_{\text{digit letter} \rightarrow \text{variability}}$  parameter was positive (0.60). Fixing this parameter to zero resulted in less, but significant, loss of fit,  $\chi^2(1, N=447)=9.0$ . Thus, the results from this model indicate that higher trial-to-trial variability reliably precedes and predicts greater decline in digit letter performance. In contrast, higher digit letter performance predicted greater subsequent 2-year increases in variability.

One important follow-up question to these findings is whether trial-to-trial variability in identical pictures performance is a more powerful predictor than the mean of identical pictures performance as observed over the same trials. To address this issue, we estimated two bivariate DCSM models, one including mean reaction time of identical pictures and categories and the other involving mean reaction time and digit letter performance. Time-to-death, chronological age, and suspected dementia were covariates. Note that mean reaction time for identical pictures was not included in the LGCM because variance in change was not reliable for this measure. However, mean reaction time can be included in a BDCSM because the estimation of the most interesting parameters (i.e., the  $\gamma$ 's) does not need variance in change (i.e., gammas and betas are fixed, not random, parameters; see Lindenberger & Ghisletta, 2004). Importantly, in neither of these models did higher mean reaction time predict steeper subsequent decline in performance. The  $-2LL$  (117,  $N=447$ ) for the model including mean reaction time and categories, which estimated 53 parameters, was 17984.4. The estimate of the inter-variable  $\gamma_{\text{identical pictures} \rightarrow \text{categories}}$  parameter was non-significant ( $-0.04$ ;  $\chi^2[1, N=447]=0.01$ ). The estimate of the inter-variable  $\gamma_{\text{categories} \rightarrow \text{identical pictures}}$  parameter was negative and significant ( $-0.92$ ;  $\chi^2[1, N=447]=17.05$ ). The  $-2LL$  (117,  $N=447$ ) for the model including mean reaction time and digit letter performance, which estimated 53 parameters, was 17327.0. The estimate of the inter-variable  $\gamma_{\text{identical pictures} \rightarrow \text{digit letter}}$  parameter was significant but positive (0.96;  $\chi^2[1, N=447]=12.83$ ), indicating that higher reaction time in identical picture lead to less decrease in digit letter performance. The estimate of the inter-variable  $\gamma_{\text{digit letter} \rightarrow \text{identical pictures}}$  parameter was negative and significant ( $-1.34$ ;  $\chi^2[1, N=447]=39.56$ ). Thus, if anything, these follow up analyses indicate that mean reaction time

Table 4

Is high trial-to-trial variability (CoV) preceding and predicting decline in mean cognitive performance?

Model	$\beta_{\text{variability}}$	$\beta_{\text{mean}}$	$\gamma_{\text{variability} \rightarrow \text{mean}}$	$\gamma_{\text{mean} \rightarrow \text{variability}}$
Variability–categories	$-0.04$ (0.22)	$-0.28$ (0.24)	$-0.84^*$ (0.30)	$-0.11$ (0.19)
Variability–digit letter	$0.57$ (0.32)	$-1.01^*$ (0.33)	$-1.30^*$ (0.44)	$0.60^*$ (0.23)
Variability–identical pictures	$-0.29$ (0.21)	$-0.58^*$ (0.28)	$0.97^*$ (0.28)	$0.31$ (0.22)

Parameter estimates for bivariate dual change score models including statistical control for the influence of time-to-death, chronological age, and suspected dementia. Note: CoV = coefficient of variation. For the estimates of the  $\gamma$  parameters, significance testing refers to significance of the  $\chi^2$  ( $df=1$ ) in nested model comparisons comparing a model fixing this parameter to 0 against the full coupling model (see Section 1.3 and Section 2). For the estimates of the  $\beta$  parameters, significance was determined from their standard errors.

\*  $p < 0.05$ .

in identical pictures is predicted by categories and digit letter performance, rather than the other way around.

To directly address the issue of the dynamic association between variability in reaction time or mean in reaction time, we estimated a BDCSM including trial-to-trial variability and the mean reaction time of identical pictures, with time-to-death, chronological age, and suspected dementia as covariates. The last row of Table 4 displays the results from this model, which estimated 53 parameters with a  $-2LL$  (117,  $N=447$ ) of 18,087.6. The estimate of the  $\gamma_{\text{variability} \rightarrow \text{identical pictures}}$  parameter was positive (0.97) and significant,  $\chi^2(1, N=447) = 19.0$ . The estimate of the  $\gamma_{\text{identical pictures} \rightarrow \text{variability}}$  parameter (0.31) was not significant,  $\chi^2(1, N=447) = 1.2$ . Thus, higher variability reliably predicts greater increases in reaction time, but not the other way around.

To summarize, higher trial-to-trial variability in identical pictures performance reliably precedes and predicts subsequent decline in categories, digit letter, and increase in mean reac-

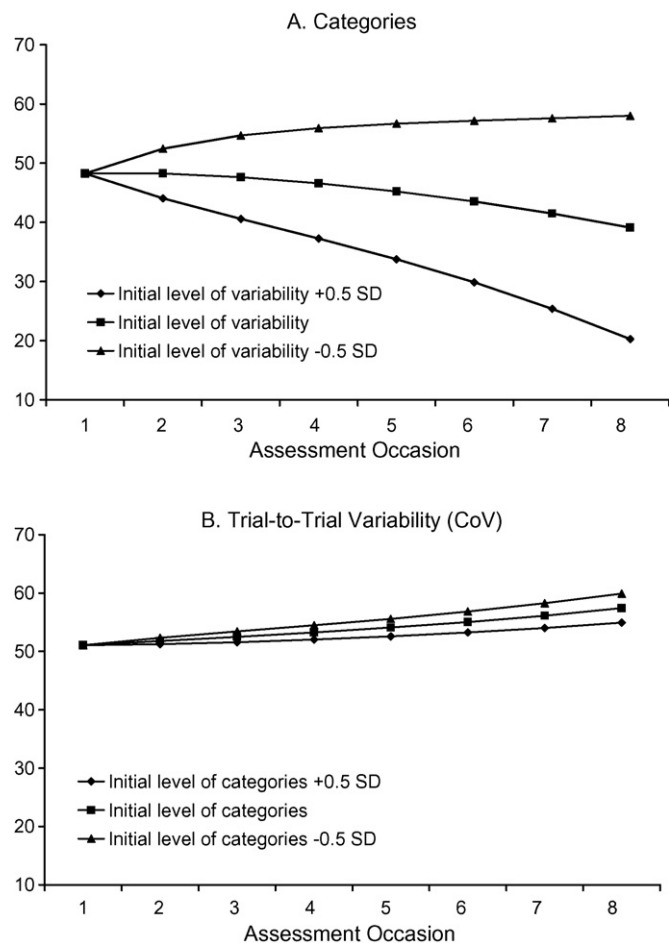


Fig. 3. Is high trial-to-trial variability (coefficient of variation, CoV) preceding and predicting decline in categories performance? Model implied means from a bivariate dual change score model (full coupling model; McArdle & Hamagami, 2001) for categories performance (A) and trial-to-trial variability (B) as function of time of assessment and varied initial (t1) means ( $-0.5SD$ , estimated mean, and  $+0.5SD$ ) of trial-to-trial variability (A) and categories (B). The figures show that lower initial trial-to-trial variability is associated with a less negative development of categories performance (A), whereas the initial levels of categories performance have little impact on development of trial-to-trial variability (B).

tion time. These effects are independent of the within-variable lead-lag effects (i.e., the  $\beta$ 's), suspected dementia status, chronological age, and time-to-death. In contrast, mean performance in identical pictures was not a potent predictor of subsequent decline in categories and digit letter.

In order to illustrate the take-home message, we first varied the initial sample means (levels) of trial-to-trial variability (CoV) by half of a standard deviation (i.e., five  $T$ -scores up and down from the estimated initial mean) while keeping initial categories performance constant. Using estimates from the full coupling model,<sup>3</sup> we then calculated the resulting implied means of categories performance. The results are displayed in Fig. 3A. As is evident in this figure, level of trial-to-trial variability has profound impact on subsequent changes in categories performance: a hypothetical individual with initial trial-to-trial variability of  $0.5SD$ s above the mean and with initial mean categories performance is expected to decline substantially more on categories performance than a hypothetical individual with lower variability.<sup>4</sup> In contrast, varying initial categories performance has negligible impact on the trajectories of variability (Fig. 3B).

### 3. Discussion

In line with MacDonald et al. (2003), we observed a strong association ( $r_s = -0.82$  and  $-0.68$ ) between longitudinal changes in intraindividual reaction time variability on a measure of perceptual speed and longitudinal changes in levels of performance in ideational fluency and perceptual speed. In addition, and more importantly, our results reveal that greater trial-to-trial variability in perceptual speed performance predicts looming changes in cognitive performance levels. In contrast, lower cognitive performance levels have negligible influences on subsequent 2-year changes in variability. These results are not confounded by within-variable dynamic effects, chronological age, suspected dementia, and time-to-death. Together, the findings substantiate the conjecture that cognitive intraindividual variability constitutes a developmentally early flag for impending old-age changes in mean levels of cognitive performance (Hultsch & MacDonald, 2004; Li & Lindenberger, 1999; Lindenberger et al., 2006; cf. Van der Maas & Molenaar, 1992).

The main finding of a dynamic influence of variability on mean cognitive performance levels is consistent with findings indicating that cognitive variability predicts chronological age (e.g., Hultsch et al., 2002), cognitive functioning (e.g., Hultsch et al., 2000), and disease status (e.g., Fuentes et al., 2001; Hogan et al., 2006; see Hultsch & MacDonald, 2004; MacDonald et al., 2006, for reviews) independently of central tendency. However, the finding is inconsistent with recent results reported by Salthouse and Berish (2005). In their cross-sectional study,

<sup>3</sup> The formula for calculating the implied means for  $x[t] > x[1]$  is:  $x[t] = 1Xs + (1 + \beta x) \times x[t - 1] + \gamma_{y-x}y[t - 1]$ , when applied for a person of average age, time-to-death, and suspected dementia status. For  $x[1]$  the mean equals  $x_0$ .

<sup>4</sup> Note that the scenario somewhat exaggerates the findings because initial level of variability and initial level of categories performance is correlated.



variance-partitioning techniques were applied to data from age-heterogeneous (from young adulthood to old age) samples. Their results showed that statistical control of mean perceptual speed performance had greater attenuating effects on correlations among variability (*SD*) in perceptual speed and cognitive performance levels than control of variability had on correlations among the mean and other cognitive performance levels. These findings were taken to indicate that, for explaining associations between age and cognitive variables, performance variability might be secondary to central tendency. The present study clearly differs from the Salthouse and Berish study in many ways. For example, the current sample is considerably older. Also, the analyses in the present study were conducted with structural equation modeling. Thus, the estimates of particular substantive interest (i.e., the  $\gamma$ 's) are attenuated for differences in reliability between variables. An alternative explanation of the Salthouse and Berish findings, as pointed out by the authors, is that variability measures have lower reliability than measures of central tendency, especially when the number of trials is small (Schmiedek, 2006). If reliability differences are left unconsidered and everything else being equal, variability measures will account for less variance in other measures than the central tendency. Moreover, the present results are based on longitudinal data, which, in contrast to cross-sectional data, can reveal dynamic associations among variables. Finally, the present study investigated variability on the trial-to-trial level in contrast to variability across closely spaced occasions. Though nested, these measures may capture different types of underlying processes.

The present findings suggest that within-person cognitive variability deserves a more prominent position both in clinical settings and in theories of neurocognitive aging. In clinical settings, cognitive variability may serve as an early warning of imminent cognitive decline and therefore carry important clinical benefits (Hultsch & MacDonald, 2004; MacDonald et al., 2006). The findings also suggest that theories of neurocognitive aging need to account for both age-related decline in mean performance and increases in cognitive variability, and need to incorporate assumptions about the developmental sequence, or cascade, of changes in variability and central tendency.

Theories highlighting the role of cognitive control and the integrity of the prefrontal cortex for age-related declines in cognitive performance (e.g., Buckner, 2004; West, 1996) have the potential to posit a developmental sequence for cognitive changes in variability and performance levels with advancing age. Marked and early structural (e.g., Raz et al., 2005) and functional (Buckner, 2004) alterations in the integrity of prefrontal cortex occur in healthy aging. Moreover, cognitive control processes supported by the prefrontal cortex are important for a wide range of cognitive functions. Importantly, cognitive variability in several different tasks might, to a high degree, reflect the integrity of cognitive control processes, such as the extent of attentional lapses and reliability of context maintenance (Bellgrove et al., 2004; Bunce et al., 1993; Hultsch et al., 2002; MacDonald et al., 2006; Stuss et al., 2003; West et al., 2002; Williams et al., 2005). Behavioral (Hultsch et al., 2002), clinical (Winterer & Weinberger, 2004), and neuroimaging (Bellgrove et al., 2004)

work support this interpretation. Thus, intraindividual cognitive variability may constitute a sensitive marker of early decline in cognitive control, which later will decline more substantially and cause widespread effects on cognitive functioning. In turn, cognitive control may become increasingly important for various types of cognitive abilities in old age, as suggested by, for example, the pattern of dedifferentiation in which various measures of cognitive performance display increasing correlations among each other (e.g., de Frias, Lövdén, Lindenberger, & Nilsson, in press; Li, Lindenberger et al., 2004). Thus, category fluency and digit letter mean performance, although perhaps best construed as markers of verbal fluency and perceptual speed, may to a greater extent depend on cognitive control in particular in old and very old age.

Theoretical models that are able to more directly relate the effects of behavioral variability on behavioral longitudinal decline to the neural level may help to promote a more mechanistic understanding of the present findings, and inform empirical neurobiological research. As a precondition for making predictions that correspond to the findings reported in this study, these models need to relate accuracy to latency, and variability to central tendency, at both behavioral and neuronal levels of analysis. In doing so, they need to recognize that variability itself, or neuronal noise, significantly contributes to the cortical dynamics of neuronal signal transmission (Stein, Gossen, & Jones, 2005). Two models, although currently not able to fully explain the current finding (e.g., the temporal ordering of the influence of variability and longitudinal cognitive declines), seem especially promising.

First, Li et al. (e.g., Li & Lindenberger, 1999; Li, Aggen et al., 2001; Li, Lindenberger et al., 2001) have proposed a neurocomputational framework that simulates senescent changes in behavior as a function of age-related reductions in the efficacy of dopaminergic neuromodulation. According to the model, deficient neuromodulation results in greater neuronal noise, and alters the extent to which neural information processing can benefit from noise tuning (e.g., stochastic resonance; Li, Brehmer et al., 2006; Li, Oertzen et al., 2006). Importantly, this model specifies a direct link between variability and central tendency in cognitive performance, as the model parameter used to simulate age changes in dopaminergic neuromodulation affects both the variability and central tendency (Li & Lindenberger, 1999; Li, Aggen et al., 2001; Li, Lindenberger et al., 2001). Given that age changes in dopaminergic neuromodulation have pronounced effects on cognitive control and fluid intelligence (cf. Li, Brehmer et al., 2006), this model may provide a viable starting point for establishing a specific link between moment-to-moment variability and subsequent longitudinal mean decline at the neuronal level of analysis. For example, with the assumption that the effects of manipulations of the gain parameter on variability is particularly tightly linked with dopaminergic modulation of the integrity of prefrontal cortex's cognitive control functions (e.g., Egan et al., 2001), temporal lead-lag association might be incorporated also within this framework.

Second, the diffusion model by Ratcliff and Rouder (1998) has been successfully used to model adult age differences in accuracy and latency of simple cognitive tasks (e.g., Ratcliff,

Thapar, Gomez, & McKoon, 2004), and has recently been extended to predict individual differences in working memory and cognitive control (e.g., Schmiedek et al., *in press*; see also Ratcliff, Schmiedek, & McKoon, *submitted for publication*). Similar to the approach taken by Li and colleagues, the diffusion model directly links variability and central tendency in cognitive performance. Taken this together with the role played by diffusion processes in modeling neural signaling (Smith & Ratcliff, 2004), the modeling approach by Ratcliff and co-workers may point to neural mechanisms that cause senescent changes in the variability of perceptual decision making, and predict longitudinal decline.

In sum, research lines directly linking intraindividual cognitive variability to brain functions at the neuroanatomical and neurochemical levels is clearly called for to improve understanding the developmental dynamics between cognitive performance and variability in old age (see also MacDonald et al., 2006). Such research might also shed light on whether intraindividual cognitive variability is best conceived of as an agent (Li & Lindenberger, 1999; Lindenberger & Oertzen, 2006) or a flag (Hultsch & MacDonald, 2004) of impending cognitive performance changes in old age.

The validity of the present findings is conditional upon several statistical assumptions, such as sample homogeneity, data missing-at-random, and the equivalence of structural relations based on interindividual and intraindividual variance. The results must also be interpreted with the timescales in mind, considering both intraindividual variability (i.e., trial-to-trial fluctuations) and intraindividual changes (i.e., 2-year changes). In other words, the present findings may not generalize to other timescales. In addition, our measure of intraindividual variability is admittedly suboptimal due to the limited amounts of trials available. However, if anything, this is likely to work against finding a dynamic influence of trial-to-trial variability on performance levels. It should also be noted that cognitive aging is clearly not fully described by a bivariate model. In other words, we may have failed to model several variables that may play important roles in the association between variability and cognitive performance. Future studies should extend directional hypotheses to higher-order multivariate models. Finally, given the high mean age of the sample, we do not know whether the findings generalize to late middle adulthood or early old age, when the first signs of longitudinal cognitive decline become manifest. Thus, the finding of developmental ordering of variability and cognitive performance levels in old age and very old age clearly deserves attempts for replication and extension.

To conclude, we found that changes in trial-to-trial variability were associated with changes in cognitive performance levels, and that higher trial-to-trial variability precedes and predicts decline in cognitive performance in old age. Trial-to-trial variability may constitute an early and clinically relevant sign of coming changes in cognitive functioning. Theories of neurocognitive aging need to account for senescent decline in both variability and central tendency, and for the developmental cascade between variability and performance.

## Acknowledgements

This study was carried out in the context of the Berlin Aging Study (BASE). The authors express their gratitude to all colleagues in BASE. Special thanks go to Florian Schmiedek for several productive discussions.

## References

- 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. (1999). Sensorimotor variables 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.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.
- Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: Smallwaters.
- Bäckman, L., Nyberg, L., Lindenberger, U., Li, S.-C., & Farde, L. (2006). The correlative triad among aging, dopamine, and cognition: Current status and future prospects. *Neuroscience and Biobehavioral Reviews, 30*, 791–807.
- Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12*, 12–21.
- Baltes, P. B., & Mayer, K. U. (1999). *The Berlin Aging Study: Aging from 70 to 100*. New York, NY: Cambridge University Press.
- Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Oxford, England: Brooks/Cole.
- Bellgrove, M. A., Hester, R., & Garavan, H. (2004). The functional neuroanatomical correlates of response variability: Evidence from a response inhibition task. *Neuropsychologia, 42*, 1910–1916.
- Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., et al. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General, 130*, 746–763.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158.
- Buckner, R. L. (2004). Memory and executive function in aging and ad: Multiple factors that cause decline and reserve factors that compensate. *Neuron, 44*, 195–208.
- 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.
- Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London, 302*, 341–359.
- 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.
- de Frias, C. M., Lövdén, M., Lindenberger, U., & Nilsson, L. -G. (in press), Schmiedek (in press). Revisiting the dedifferentiation hypothesis with longitudinal multi-cohort data. *Intelligence*.
- Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., et al. (2001). Effect of COMT val108/158 met genotype on frontal lobe function and risk for schizophrenia. In *Proceedings of the National Academy of Sciences* 98 (pp. 6917–6922).
- Eizenman, D. R., Nesselroade, J. R., Featherman, D. L., & Rowe, J. W. (1997). Intraindividual variability in perceived control in an older sample: The MacArthur successful aging studies. *Psychology and Aging, 12*, 489–502.
- Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for kit of factor-referenced cognitive tests*. NJ: Educational Testing Service.

- Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology*, *40*, 935–952.
- Fozard, J. L., Verduyssen, M., Reynolds, S. L., Hancock, P., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore longitudinal study of aging. *Journals of Gerontology*, *49*, 179–189.
- Fuentes, K., Hunter, M. A., Strauss, E., & Hultsch, D. F. (2001). Intraindividual variability in cognitive performance in persons with chronic fatigue syndrome. *Clinical Neuropsychologist*, *15*, 210–227.
- Gerstorff, D., Herlitz, A., & Smith, J. (2006). Stability of cognitive sex differences in advanced old age: The role of education and attrition. *Journal of Gerontology: Psychological Sciences*, *61*, 245–249.
- Gerstorff, D., Lövdén, M., Röcke, C., Smith, J., & Lindenberger, U. (2007). Well-being affects changes in perceptual speed in advanced old age: Longitudinal evidence for a dynamic link. *Developmental Psychology*, *43*, 705–718.
- Ghisletta, P. (2005). *Power to identify lead-lag relationships in bidimensional dynamic systems based on latent difference score models presented at the meeting of the Gerontological Society America*. USA: Orlando.
- Ghisletta, P., & de Ribaupierre, A. (2005). A dynamic investigation of cognitive dedifferentiation with control for retest: Evidence from the swiss interdisciplinary longitudinal study on the oldest old. *Psychology and Aging*, *20*, 671–682.
- Ghisletta, P., & Lindenberger, U. (2003). Age-based structural dynamics between perceptual speed and knowledge in the Berlin Aging Study: Direct evidence for ability dedifferentiation in old age. *Psychology and Aging*, *18*, 696–713.
- Ghisletta, P., & Lindenberger, U. (2005). Exploring structural dynamics within and between sensory and intellectual functioning in old and very old age: Longitudinal evidence from the Berlin Aging Study. *Intelligence*, *33*, 555–587.
- Ghisletta, P., Bickel, J.-F., & Lövdén, M. (2006). Does activity engagement protect against cognitive decline in old age? Methodological and analytical considerations. *Journal of Gerontology: Psychological Sciences*, *61*, 253–261.
- Helmchen, H., Baltes, M. M., Geiselmann, B., Kanowski, S., Linden, M., & Reischies, F. M. (1999). Psychiatric illnesses in old age. In P. B. Baltes & K. U. Mayer (Eds.), *The Berlin Aging Study: Aging from 700 to 100* (pp. 167–196). New York, NY: Cambridge University press.
- Hendrickson, A. E. (1982). The biological basis of intelligence: I. Theory. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 151–196). Berlin, Germany: Springer Verlag.
- Hogan, M. J., Carolan, L., Roche, R. A. P., Dockree, P. M., Kaiser, J., Bunting, B. P., Robertson, I. H., & Lawlor, B. A. (2006). Electrophysiological and information processing variability predicts memory decrements associated with normal age-related cognitive decline and Alzheimer's disease (AD). *Brain research*, *1119*, 215–226.
- 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, UK: Oxford University Press.
- Hultsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *57*, 101–115.
- Hultsch, D. F., MacDonald, S. W., 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, Germany: Springer Verlag.
- Klein, L. E., Roca, R. P., McArthur, J., Vogelsang, G., Klein, G. B., & Kirby, S. M. (1985). Diagnosing dementia. Univariate and multivariate analyses of the mental status examination. *Journal of the American Geriatrics Society*, *33*, 483–488.
- Li, S.-C. (2005). Neurocomputational perspectives linking neuromodulation, processing noise, representational distinctiveness, and cognitive aging. In R. Cabeza, L. Nyberg, & D. Park (Eds.), *Cognitive neuroscience of aging: Linking cognitive and cerebral aging* (pp. 354–379).
- Li, S.-C., & Lindenberger, U. (1999). Cross-level unification: A computational exploration of the link between deterioration of neurotransmitter systems and dedifferentiation of cognitive abilities in old age. In L.-G. Nilsson & H. J. Markowitsch (Eds.), *Cognitive neuroscience of memory* (pp. 103–146). Kirkland: Hogrefe & Huber Publishers.
- Li, S.-C., & Sikstrom, S. (2002). Integrative neurocomputational perspectives on cognitive aging, neuromodulation, and representation. *Neuroscience & Biobehavioral Reviews*, *26*, 795–808.
- Li, S.-C., Aggen, S. H., Nesselroade, J. R., & Baltes, P. B. (2001). Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The MacArthur successful aging studies. *Gerontology*, *47*, 100–116.
- Li, S.-C., Brehmer, Y., Shing, Y. L., Werkle-Bergner, M., & Lindenberger, U. (2006). Neuromodulation of associative and organizational plasticity across the life span: Empirical evidence and neurocomputational modeling. *Neuroscience and Biobehavioral Reviews*, *30*, 775–790.
- 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., & Frensch, P. A. (2000). Unifying cognitive aging: From neuromodulation to representation to cognition. *Neurocomputing: An International Journal*, *32-33*, 879–890.
- 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., & Sikstrom, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences*, *5*, 479–486.
- Li, S.-C., Naveh-Benjamin, M., & Lindenberger, U. (2005). Aging neuromodulation impairs associative binding: A neurocomputational account. *Psychological Science*, *16*, 445–450.
- Li, S.-C., Oertzen, T. v., & Lindenberger, U. (2006). A neurocomputational model of stochastic resonance and aging. *Neurocomputing*, *69*, 1553–1560.
- Lindenberger, U., & Ghisletta, P. (2004). Modeling longitudinal changes in old age: From covariance structures to dynamic systems. In R. A. Dixon, L. Bäckman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 199–216). Oxford: Oxford University Press.
- Lindenberger, U., & Oertzen, T. v. (2006). Variability in cognitive aging: From taxonomy to theory. In F. I. M. Craik & E. Bialystok (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 297–314). Oxford, UK: Oxford University Press.
- Lindenberger, U., Gilberg, R., Little, T. D., Nuthman, R., Pötter, U., & Baltes, P. B. (1999). Sample selectivity and generalizability of the results of the Berlin Aging Study. In P. B. Baltes & K. U. Mayer (Eds.), *The Berlin Aging Study: Aging from 70 to 100* (pp. 56–83). Cambridge, England: Cambridge University Press.
- Lindenberger, U., Li, S.-C., & Bäckman, L. (2006). Delineating brain-behavior mappings across the lifespan: Substantive and methodological advances in developmental neuroscience. *Neuroscience and Biobehavioral Reviews*, *30*, 713–717.
- Lindenberger, U., Mayr, U., & Kliegl, R. (1993). Speed and intelligence in old age. *Psychology and Aging*, *8*, 207–220.
- Lindenberger, U., Singer, T., & Baltes, P. B. (2002). Longitudinal selectivity in aging populations: Separating mortality-associated versus experimental components in the Berlin Aging Study (BASE). *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *57*, 474–482.
- Lövdén, M., Ghisletta, P., & Lindenberger, U. (2004). Cognition in the Berlin Aging Study (BASE): The first 10 years. *Aging, Neuropsychology, and Cognition*, *11*, 104–133.
- Lövdén, M., Ghisletta, P., & Lindenberger, U. (2005). Social participation attenuates decline in perceptual speed in old and very old age. *Psychology and Aging*, *20*, 423–434.
- 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*, 474–480.

- Mäntylä, T., & Bäckman, L. (1990). Encoding variability and age-related retrieval failures. *Psychology and Aging*, 5, 545–550.
- McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 137–176). Washington, DC: American Psychological Association.
- McArdle, J. J., & Nesselroade, J. R. (1994). Using multivariate data to structure developmental change. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological contributions* (pp. 223–267). Hillsdale, NJ: Erlbaum.
- McArdle, J. J., Hamagami, F., Jones, K., Jolesz, F., Kikinis, R., Spiro, A., III, et al. (2004). Structural modeling of dynamic changes in memory and brain structure using longitudinal data from the normative aging study. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 59, 294–304.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- 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., & Boker, S. M. (1994). Assessing constancy and change. In T. Brotherton & J. Weinberger (Eds.), *Can personality change?* (pp. 121–147). Washington, DC: American Psychological Association.
- Park, D. C., Polk, T. A., Park, R., Minear, M., Savage, A., & Smith, M. R. (2004). Aging reduces neural specialization in ventral visual cortex. *PNAS*, 101, 13091–13095.
- 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. M. A., 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 A: Human Experimental Psychology*, 54, 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., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356.
- Ratcliff, R., Schmiedek, F., & McKoon, G. A. (submitted for publication). Diffusion model explanation of the worst performance rule for reaction time and IQ.
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology & Aging*, 19, 278–289.
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., et al. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, 15, 1679–1689.
- Robertson, S., Myerson, J., & Hale, S. (2006). Are there age differences in intraindividual variability in working memory performance? *Journal of Gerontology: Psychological Sciences*, 61, 18–24.
- Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, 88, 245–258.
- Rubin, D. (1974). Characterizing the estimation of parameters in incomplete-data problems. *Journal of American Statistical Association*, 69, 467–474.
- Salthouse, T. A. (1993). Attentional blocks are not responsible for age-related slowing. *Journals of Gerontology*, 48, 263–270.
- Salthouse, T. A., & Berish, D. E. (2005). Correlates of within-person (across-occasion) variability in reaction time. *Neuropsychology*, 19, 77–87.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological methods*, 7, 147–177.
- Schmiedek, F. (2006). *Overview talk for the plenary session on intraindividual differences and change presented at the cognitive aging conference.*
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. -M., & Wittman, W. W. (in press). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*.
- 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 in Psychological Science*, 3, 1–5.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27, 161–168.
- 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.
- Stein, R. B., Gossen, E. R., & Jones, K. E. (2005). Neuronal variability: Noise or a part of the signal. *Nature Reviews Neuroscience*, 6, 389–397.
- Strauss, E., MacDonald, S. W., Hunter, M., Moll, A., & Hultsch, D. F. (2002). Intraindividual variability in cognitive performance in three groups of older adults: Cross-domain links to physical status and self-perceived affect and beliefs. *Journal of the International Neuropsychological Society*, 8, 893–906.
- Stuss, D. T., Murphy, K. J., Binns, M. A., & Alexander, M. P. (2003). Staying on the job: The frontal lobes control individual performance variability. *Brain: A Journal of Neurology*, 126, 2363–2380.
- 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.
- Sullivan, E. V., & Pfefferbaum, A. (2006). Diffusion tensor imaging and aging. *Neuroscience and Biobehavioral Reviews*, 30, 749–761.
- Thaler, D. S. (2002). Design for an aging brain. *Neurobiology of Aging*, 23, 13–15.
- Van der Maas, H. L., & Molenaar, P. C. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review*, 99, 395–417.
- Welford, A. T. (1965). Performance, biological mechanisms, and age: A theoretical sketch. In A. T. Welford & J. E. Birren (Eds.), *Behavior, aging, and the nervous system* (pp. 3–20). Springfield, IL: Thomas.
- West, R. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, 120, 272–292.
- West, R., Murphy, K. J., Armilio, M. L., Craik, F. I., & 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.
- Winterer, G., & Weinberger, D. R. (2004). Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends in Neurosciences*, 27, 683–690.