Does variability in cognitive performance correlate with frontal brain volume?

Martin Lövdén [a,b,c,*], Florian Schmiedek [a,d], Kristen M. Kennedy [e], Karen M. Rodrigue [e], Ulman Lindenberger [a], Naftali Raz [a,f]

[a] Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany
[b] Aging Research Center, Karolinska Institutet & Stockholm University, Stockholm, Sweden
[c] Department of Psychology, Lund University, Sweden
[d] German Center for International Educational Research (DIPF), Frankfurt am Main, Germany
[e] School of Behavioral and Brain Sciences, Center for Vital Longevity, University of Texas at Dallas, Dallas, USA
[f] Department of Psychology, Wayne State University, Detroit, USA

ARTICLE INFO

ABSTRACT

Little is known about the neural correlates of within-person variability in cognitive performance. We investigated associations between regional brain volumes and trial-to-trial, block-to-block, and day-to-day variability in choice–reaction time, and episodic and working memory accuracy. Healthy younger ($n = 25$) and older ($n = 18$) adults underwent 101 daily assessments of cognitive performance, and their regional brain volumes were measured manually on magnetic resonance images. Results showed that smaller prefrontal white matter volumes were associated with higher block-to-block variability in choice–reaction time performance, with a stronger association observed among older adults. Smaller volumes of the dorsolateral prefrontal cortex covaried with higher block-to-block variability in episodic memory (number–word pair) performance. This association was stronger for younger adults. The observed associations between variability and brain volume were not due to individual differences in mean performance. Trial-to-trial and day-to-day variability in cognitive performance were unrelated to regional brain volume. We thus report novel findings demonstrating that block-by-block variability in cognitive performance is associated with integrity of the prefrontal regions and that between-person differences in different measures of variability of cognitive performance reflect different age-related constellations of behavioral and neural antecedents.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Much attention has been paid to individual differences in cognitive performance and age-related differences therein (Craik and Salthouse, 2008). However, less is known about intra-individual variability in cognitive performance, which is defined as lawful but transient within-person changes in behavior within a defined period (Nesselroade, 1991) and which can be observed across a wide range of tasks (Hultsch et al., 2008; MacDonald et al., 2006; Rabbitt et al., 2001). Our ignorance is even greater when it comes to the neural underpinnings of intra-individual variability in cognitive performance (MacDonald et al., 2006, 2009b).

For all its apparent simplicity, the concept of intra-individual variability in cognitive performance can be approached in many meaningful ways. Speed and accuracy characteristics of performance may fluctuate on a yearly, monthly, daily, hourly, or momentary basis, thus demanding examination on various scales, with a wide range of

resolution (Boker et al., 2009). Moreover, selection of scale is not just a decision about measurement as, depending on the time scale, variability may reflect different behavioral and neural antecedents (Boker et al., 2009; Lindenberger and von Oertzen, 2006; MacDonald et al., 2009b; Rabbitt et al., 2001). For example, trial-to-trial variability in performance on simple reaction time (RT) tasks may reflect lapses of attention (Bunce et al., 1993; West et al., 2002; Williams et al., 2005) and admixture of neural noise to task-related processes (Bäckman et al., 2006; Li et al., 2001; MacDonald et al., 2009a). Such factors may also induce variability in performance across blocks of trials, but block-to-block variability may additionally reflect influences such as search and selection of optimal strategies (Allaire and Marsiske, 2005; Li et al., 2004; Shing et al., 2012; Siegler, 1994). Day-to-day variability in cognitive performance may stem from fluctuations in non-cognitive factors, such as stress (Sliwinski et al., 2006) or motivation (Brose et al., 2010). Tasks may also vary in the degree to which they tap different antecedents of variability in cognitive performance. Simple RT tasks present relatively little opportunity for exploration of strategies. In contrast, working memory (e.g., Shing et al., 2012) and episodic memory (e.g., Kirchhoff, 2009) tasks provide ample room for within-subject differences in strategies, which may contribute to performance variability.

* Corresponding author at: Aging Research Center, Karolinska Institutet & Stockholm University, Gävlegatan 16, 113 30 Stockholm, Sweden.
E-mail address: Martin.Lovden@ki.se (M. Lövdén).

In sum, intra-individual variability in cognitive performance is not a unitary construct (Allaire and Marsiske, 2005; Boker et al., 2009; Lindenberger and von Oertzen, 2006; MacDonald et al., 2009b), and its heterogeneity should inform the search for neural correlates of intra-individual variability. However, extant studies on brain correlates of variability in cognitive performance have focused almost exclusively on fluctuations in response latency across trials of simple and choice RT tasks. These studies report that higher trial-to-trial variability in RT is associated with smaller volume (Anstey et al., 2007; Walhovd and Fjell, 2007) and lower integrity of white matter in frontal, parietal, temporal, and central brain regions (Bunce et al., 2007; Fjell et al., 2011; Moy et al., 2011; Tamnes et al., 2012). In a related vein, variability in finger tapping is associated with white matter integrity (Ullén et al., 2008). Although at least two studies (Moy et al., 2011; Ullén et al., 2008) failed to observe significant associations between gray matter volume and variability, other research suggests that trial-to-trial variability may be linked to prefrontal cortex integrity. For example, individuals with lesions in the dorsolateral prefrontal cortices (DLPFC) display greater variability in performance on simple and complex RT tasks than patients with orbitofrontal or non-frontal lesions and healthy controls (Stuss et al., 2003). In primates, direct pharmacological manipulations of the DLPFC affect trial-to-trial variability in saccade response time (Pouget et al., 2009). In addition, individual differences in cognitive variability have been linked to activation in the attentional control network, including the prefrontal cortices, in a number of functional imaging studies (Bellgrove et al., 2004; Prado et al., 2011; Simmonds et al., 2007).

The extant literature contains no studies of neural correlates of block-to-block variability in the accuracy of performing complex tasks geared to assess working memory and episodic memory. It is therefore unknown whether the association between white matter integrity and variability is unique for RT tasks and to the time-scale of trial-to-trial measurements, or whether the observed associations generalize to other measures of variability extracted over different time scales and with indices unrelated to speed of response. We also note that slower performance fluctuations, similar to block-to-block variability, are confounding many measures of trial-to-trial variability used in previous reports. Hence, there is a need for a more systematic and nuanced investigation of the neural underpinnings of variability in cognitive performance.

In this study, we address the outlined lacunae in knowledge by taking advantage of a unique data set from the COGITO study (Schmiedek et al., 2010b), an investigation of younger and older adults who underwent 101 daily assessments of cognitive performance. The COGITO study presents an opportunity to follow methodological recommendations (e.g., Boker et al., 2009), and simultaneously examine and directly compare intra-individual and inter-individual variability on three nested time scales: trial-to-trial, block-to-block, and day-to-day. In the COGITO study, we assessed cognitive performance at various levels of complexity, ranging from choice–reaction time (CRT) to working and episodic memory accuracy. The CRT task allowed for estimation of all three levels of variability, while the accuracy tasks allow for estimation of block-to-block and day-to-day variability. To examine neural substrates of cognitive variability, we measured volumes in several brain regions that differed in their putative relevance to the cognitive tasks.

We based our prediction on the empirical review as well as the theoretical link between prefrontal cortex function and transient failures of cognitive control (Miller and Cohen, 2001; Weissman et al., 2006), which have been related to trial-to-trial variability (e.g., West et al., 2002; Williams et al., 2005). In addition to the reviewed evidence, findings link white matter integrity to speeded performance (e.g., Bender and Raz, 2012; Bucur et al., 2008; Espeseth et al., 2006; Kennedy and Raz, 2009) suggesting that white rather than gray matter of the PFC would be relevant to variations in speed. We thus hypothesized that trial-to-trial and block-to-block variability in speed-based indices of CRT performance would be associated specifically with the prefrontal white matter volume. In addition, we predicted that greater trial-to-trial and block-to-block variability in accuracy of cognitive performance would be associated with smaller frontal gray matter volumes, but that no associations between day-to-day variability in cognitive performance and brain volume would emerge. As the association between trial-to-trial variability in RT and white matter integrity has been reported to increase in old age (Fjell et al., 2011), this association may emerge only in old age due to correlated individual differences in decline of variability and white matter integrity. Thus, we predicted such an increase with age group for trial-to-trial and block-to-block variability of CRT performance. We assumed that block-to-block variability in working and episodic memory accuracy might contain larger influence from search and selection of optimal strategies. We expected that such adaptive influences on variability of cognitive performance might mask the predicted age-related increase in the association between variability of cognitive performance and brain volume.

Material and methods

Participants

Participants were recruited through newspaper advertisements, word-of-mouth recommendations, and fliers circulated in Berlin, Germany (see Schmiedek et al., 2010b for details). The main COGITO study involved 101 younger (aged 20–31 years) and 103 older adults (aged 65–80 years). Out of these participants, 30 younger and 27 older individuals volunteered and were eligible for magnetic resonance imaging (MRI). To be eligible, participants had to report being right-handed, have normal or corrected-to-normal vision, and report no cardiovascular disease (except treated hypertension, recorded in seven older adults), diabetes, neurological or psychiatric conditions, use of anti-seizure or antidepressant drugs, or drug or alcohol abuse. Scores on the Mini-Mental State Examination (Folstein et al., 1975) were all above 25. Based on evaluations of the anatomical images by a clinical neurologist, one younger and six older adults were excluded due to various brain abnormalities, and one older participant was excluded from this study because of significant movement artifacts. Four younger and two older participants dropped out during the course of the study.

Thus, the sample with complete data consisted of 25 younger (13 women/12 men) and 18 older (9 women/9 men) adults. Descriptive statistics at pretest (see Table 1) showed that the sample displayed the typical age-related differences: lower perceptual speed (Digit-symbol substitution; Wechsler, 1981) and higher vocabulary (Spot-a-word; Lindenberger et al., 1993) scores in the older participants. The age groups were comparable on self-reported years of education. Relative to the mean Digit-symbol scores reported in a meta-analysis (Hoyer et al., 2004), the sample was less positively selected (about 1.1 SD for the younger and 0.6 SD for the older group) than typical samples in cognitive aging research.

Table 1
Participant characteristics at pretest.

Measure	Younger		Older		$F(1, 42)$	p
	M	SD	M	SD		
Age	25.0	3.2	70.1	3.8	1793.8	<.001
Digit-symbol	59.8	9.3	44.6	6.9	34.3	<.001
Spot-a-word	0.66	0.11	0.82	0.07	32.5	<.001
Years of education	16.9	3.1	16.3	3.9	0.3	.572
Daily sessions	101.8	3.1	100.0	3.8	3.0	.089

The F- and p-values correspond to the age effect in a one-way ANOVA. Digit-symbol — Digit-Symbol Substitution Test (Wechsler, 1981); Spot-a-word — a German vocabulary test (Lindenberger et al., 1993); daily sessions — the number of sessions completed during the longitudinal phase of the study.

Participants received an honorarium between 1450 and 1950 Euro for the total project participation, depending on the number of completed sessions and their pace of completing the longitudinal phase of the study. The ethical review board of the Max Planck Institute for Human Development, Berlin, Germany, approved the behavioral parent study, and the review board of the Otto-von-Guericke University of Magdeburg, Germany, approved the imaging study. Written informed consent was obtained prior to the investigation.

Procedures

The general design of the study included an extensive pretest assessment, an average of 101 daily assessments of cognitive performance, and an extensive posttest assessment (see Schmiedek et al., 2010a, 2010b for a detailed description).

Pretest–posttest assessment

Participants underwent 10 days of behavioral pre-testing in group sessions that lasted 2–2.5 h. Measurements consisted of self-report questionnaires, cognitive tasks included in the daily phase, and a number of covariates and transfer tasks (for a detailed description, see Schmiedek et al., 2010b). The 10 posttest group sessions (1.5–2 h each) consisted of re-administration of the pretest cognitive tasks and additional self-report measures.

The pretest brain-imaging session was conducted after the behavioral pretest and immediately before the daily assessment phase; the posttest images were acquired shortly after the completion of the behavioral posttest. Magnetic resonance images were acquired using a GE Signa LX 1.5 T system (General Electric, Milwaukee, WI) with actively shielded magnetic field gradients (maximum amplitude 40 mTm^{-1}). The MRI protocol included a T1-weighted sagittal 3D scan (contrast-optimized spoiled gradient-echo sequence, 124 slices, slice thickness = 1.5 mm, FOV 250×250 mm^2; 256×256 matrix; TE = 8 ms; TR = 24 ms; flip angle = 30°). In addition, diffusion-weighted images were collected and functional MRI images were acquired during the performance of one perceptual speed, one working memory, and one episodic memory task. Only data from the T1-weighted images are reported here.

Daily assessment phase

Participants practiced daily for 1 h on individual computerized testing stations, up to six participants per lab room. The task battery consisted of three working memory, three episodic memory, and six perceptual speed tasks. Sessions were scheduled on an individual basis on up to 6 days a week. At the end of each session, participants received feedback on their performance on all tasks, including average accuracies and reaction times.

The practiced perceptual speed tasks were three CRT tasks (odd vs. even numbers; consonants vs. vowels; symmetric vs. asymmetric figures) and three comparison tasks (two strings of digits; two strings of consonants; two three-dimensional figures). In the episodic memory tasks, participants had to memorize word lists, number–word pairs, or object positions in a grid. Working memory tasks were adapted versions of the alpha span, numerical memory updating, and spatial n-back tasks. To reduce the number of tasks in this study, and thus to avoid the need for a conservative alpha level correction that may compromise statistical power, while still maintaining heterogeneity of the abilities tapped by the task, we restricted the analyses to tasks with numerical content. That is, we focused on the numerical CRT task (odd vs. even numbers), the numerical working memory task (memory updating), and the numerical episodic memory task (number–word pairs; for details regarding the other practiced tasks see Schmiedek et al., 2010a, 2010b).

Stimuli presentation times varied during the pretest administration, thus presenting an opportunity to estimate time-accuracy functions, which describe a person's accuracy as a function of presentation time (PT). To optimize the cognitive challenge of these tasks across individuals, while maintaining motivation, the time–accuracy information was used for tailoring the difficulty of the subsequently practiced tasks by adjusting presentation time for each individual from the start of the micro-longitudinal practice phase. Specifically, for each memory task and each individual, exponential time-accuracy functions were fitted to mean accuracies for the different PT conditions at pretest. The functions included three freely estimated parameters, for onset, rate, and asymptote, as well as a lower asymptote parameter fixed to different values for each task (.10 for memory updating and .00 for the number–word pair task). The fitted values from these functions were used to choose PTs that were clearly above random guessing but below the upper level of performance, which was defined by the midpoint between the lower asymptote level and perfect accuracy (e.g., (.10 + 1.0)/2 = .55 for memory updating and .50 for number–word pairs). The minimum level was defined by the midpoint between lower asymptote level and the upper level (e.g., (.10 + .55)/2 = .325 for memory updating and .25 for number–word pairs). The PT was chosen so that the predicted performance level based on the time–accuracy function was between the minimum and the upper levels. If performance was above the upper level for the second-fastest PT, then the fastest PT was chosen, even if predicted accuracy was below the minimum level for the fastest PT. For the CRT, a similar procedure was used to individually define fast and slow masking times for each participant. The fast masking time was chosen based on a medium accuracy level of .625 and an upper level of .75, while for the slow masking time, those levels were .875 and .95, respectively. PTs and masking times were kept constant over the daily assessment phase.

In the CRT task, the stimuli were seven lines of the number "8" displayed as on a pocket calculator. Stimuli were masked with a stimulus that combined this "calculator 8" with extending lines in all 10 possible directions. Possible masking times were 1, 2, 4, or 8 screen cycles (12, 24, 47, or 94 ms). Depending on pretest performance, two of the masking times (for fast and slow conditions) were chosen for each participant. Stimuli were "3", "5", and "7", for the odd and "2", "4", and "6" for the even condition. Participants responded by button presses, indicating as fast as possible whether the stimulus was odd or even. The CRT task consisted of two blocks, which each consisted of 40 stimuli, 20 for the fast and 20 for the slow condition, with randomly chosen stimuli out of the two possible response categories. The basic unit of the dependent variable from this task was the RT to an individual trial. For this study, we only included the individual's slowest masking condition as the daily presentation time that we used to compute variability indices, in order to be consistent with past reports of trial-to-trial variability that have used high-accuracy tasks. Mean accuracy for the CRT task was .90 (SD = .04) at pretest and .91 (SD = .04) at posttest.

In the memory-updating task, four single digits (in the range of 0 to 9) were presented simultaneously for 4000 ms in each of four horizontally placed cells. After an inter-stimulus interval of 500 ms, eight updating operations were presented sequentially in a second row at varying horizontal positions. The requested updating operations were additions and subtractions in the range of −8 to +8. Those updating operations had to be applied to the memorized digits from the corresponding cells above and the updated results had to be memorized. Possible presentation times, based on pretest performance, were 500, 1250, and 2750 ms. Inter-stimulus interval was 250 ms. At the end of each trial, the four end results had to be entered in the four cells in the upper row. Eight blocks were included in each daily session. The basic unit of the dependent variable from this task was the number of correctly retrieved numbers in a block.

In the number–word pair task, lists of 12 two-digit numbers (randomly drawn) and nouns (randomly drawn from a pool of plural nouns) in plural case pairs (e.g., "23 dogs") were presented sequentially with presentation times individually adjusted based on pretest performance. Possible PTs were 1000, 2000, or 4000 ms. Inter-stimulus interval was 1000 ms. After presentation, the nouns appeared in random

order and the corresponding numbers had to be entered. Two blocks were included in each daily session. The basic unit of the dependent variable from this task was the number of correctly retrieved numbers in a block.

Data pre-processing

Behavioral data

To separate intra-individual variability from practice-induced improvements, the data for memory-updating and the number–word pair tasks were de-trended using a logistic function. For RTs in the CRT task, the exponential function served this purpose. The functions were chosen because they are frequently used in the literature to describe individual learning curves and because graphical inspections of model fit for each participant indicated that they describe individual trends well. The response data from RT in the CRT task were truncated at 200 ms and 2500 ms.

Day-to-day variability and block-to-block variability on the episodic and working memory tasks were separated using multilevel variance component estimation. For the CRT task, block-to-block variability was additionally separated from trial-to-trial variability using the same approach. This was necessary because observed variability at higher levels (e.g., day-to-day variability) partly contains variability at lower levels. For example, even if there were no systematic performance fluctuations from day to day, the presence of block-to-block variability would lead to day-to-day fluctuations of the daily means (cf. Rabbitt et al., 2001). Multilevel models allow estimating the independent contributions of variability at the trial-to-trial, block-to-block, and the day-to-day level to overall observed day-to-day variability. Variance components for trial-to-trial, block-to-block, and day-to-day variability were estimated separately for each participant using SAS PROC MIXED models with a 3-level hierarchically nested structure (i.e., containing random effects for days, blocks nested within days, and trials nested within blocks; the SAS code is available from the authors upon request). The resulting three variance parameters add up to the total variance of each participant and can therefore be compared across participants in the same way as observed variances.

Note that due to this variance partitioning approach, the day-to-day variability indices for the three tasks reflect only the portion of observed day-to-day variability that denotes systematic changes of average performance level from one day to another. In addition, block-to-block variability on the CRT task reflects only systematic changes in average performance level from one block to another and is not confounded by trial-to-trial variability, which is separately estimated for this task. The block-to-block variability indices for the two accuracy tasks (memory updating and episodic memory) may, however, partly consist of trial-to-trial variability, because this level of performance variability is not available for these tasks.

To alleviate the skew in the distributions, all day-to-day variability indices, as well as the block-to-block and trial-to-trial variability measures from the CRT task, were log-transformed. After these transformations, the distributions for all seven variability indices were accepted as normal (skewness = −0.92–2.06; kurtosis = −0.70–5.47).

For memory updating, pretest and posttest mean performance was indexed by averaging accuracy from the 750 ms and 1500 ms presentation times as all but one subject were assigned one of either PT for the micro-longitudinal assessment. For number–word pairs, mean performance was computed by averaging accuracy from the 1000 ms, 2000 ms, and 4000 ms PTs. For the CRT task, mean performance was based on the mean RT of the trials with masking times of 2, 4, and 8 cycles. The masking condition with 1 cycle was not included because only young adults received this condition. The distributions were acceptable for all of these measures (skewness = 0.06–0.97; kurtosis = −1.30–0.46).

To investigate whether the PTs assigned to participants for their daily task completion affected the estimates of variability, we analyzed the variability measures with one-way ANCOVAs with PT as the factor separately for the three tasks. For the memory updating task, a single subject receiving a PT of 3000 ms was included in the group of participants receiving a time of 1500 ms. The threshold for statistical significance was .05. Pretest performance was included as a covariate in order to investigate effects of PT on the variability measures over and beyond effects of performance level, which was the determining factor for assigning the PTs. None of the variability indices varied significantly as a function of PT, $Fs < 2.41$.

Segmentation of regional brain volume

Image processing and manual regional volumetry are described in detail elsewhere (Raz et al., 2004, 2005). The baseline and follow-up images were coded, mixed, and assigned randomly to two tracers, who were blind to the time of acquisition and to the demographic characteristics of the participants. Reliability of all ROI measures (ICC [2], intraclass correlation for random raters; Shrout and Fleiss, 1979) exceeded .93. The volumes were computed from measured areas of the ROIs and corrected for intracranial vault (ICV) size with an ANCOVA approach (see Raz et al., 2004 for details). In addition to the ICV, the following seven ROIs were measured: dorsolateral prefrontal cortex (DLPFC), orbital frontal cortex (OF), the adjacent prefrontal white matter (FW), primary visual (calcarine) cortex (VC), the hippocampus (HC), the caudate nucleus (CD), and the cerebellar hemispheres (CB).

All estimates of regional brain volume displayed a high correlation between pretest and posttest measurements (*mean r* = .97; *range* = .90–.99), indicating very high stability of individual differences over time. For examining associations with cognitive variability, we therefore averaged these measures over pretest and posttest. Because we had no hypotheses about hemispheric asymmetries, we also summed the measures across the left and the right hemispheres. The variables displayed acceptable skewness (−0.01–0.35) and kurtosis (−1.11 to –0.10).

Statistical analyses

We used a regression approach to test our hypotheses of volume–variability associations. First, we computed the incremental addition of variance in the cognitive variability indices (R^2 change) explained by regional brain volumes, after accounting for the influence of chronological age. To balance the risk for Type I and Type II errors, we used a threshold for statistical significance at .0125 (.05/4 measures of variability) for the predicted associations between block-to-block variability in the three tasks and trial-to-trial variability in the CRT task and DLPFC and FW volume. We set the correction for multiple comparisons for the rest of the analyzed association at $p = .007$ (.05/7 measures). To examine age differences in the detected associations between frontal brain volumes and block-to-block variability, we next added the volume × age interaction term to the regression equation, after entering age and brain volume. Finally, to address whether mean performance accounted for the observed associations between variability and brain volume we computed the incremental addition of variance in the cognitive variability indices explained by regional brain volumes, after accounting for the influence of chronological age and mean performance at pretest and posttest.

Results

Regional brain volumes and variability in cognitive performance

Higher block-to-block variability in CRT was associated with smaller FW volumes, $R^2_{Change} = 0.147$; $F_{Change} (1, 40) = 6.96$, $p = .012$. In addition, higher block-to-block variability in number–word pair accuracy was associated with smaller DLPFC volume, $R^2_{Change} = 0.161$; $F_{Change} (1, 40) = 16.61$, $p < .001$. The three measures of day-to-day variability in performance and the measure of trial-to-trial variability in CRT

were not significantly associated with any measure of regional brain volume,[1] all F_{Change} (1, 40) < 3.20, all ps > .077.

Age differences in association between variability and brain volume

For the measure of block-to-block variability in CRT, addition of the interaction between FW volume and age produced a significant increment in explained variance; that is, the association between cognitive variability and prefrontal white matter volume differed between the age groups, $R^2_{Change} = 0.087$; F_{Change} (1, 39) = 4.46, $p = .041$. As evident in the scatter plots displayed in Fig. 1, the association was stronger in the older ($r = -.64$, $p = .004$) than in the younger age group ($r = -.16$, $p = .440$). The pattern was the opposite for the association between DLPFC volume and block-to-block variability on the number–word task, with younger adults showing a significantly stronger association ($r = -.70$, $p < .001$) than the older group ($r = -.15$, $p = .551$), $R^2_{Change} = 0.038$; F_{Change} (1, 39) = 4.19, $p < .047$.

To probe the robustness of the observed associations, we performed bootstrap analyses (10,000 bootstraps) on the correlation coefficient for the association between block-to-block variability in CRT and FW volume in the older adults and the correlation coefficient for the association between block-to-block variability on the number–word task and DLPFC volume in the younger adults. These analyses confirmed the significance tests reported above: The 95% confidence intervals (estimated with the accelerated bias-corrected percentile method) indicated significant associations between block-to-block variability in CRT and FW volume in the older adults (confidence interval = −.84 to −.22) and between block-to-block variability on the number–word task and DLPFC volume in the younger adults (confidence interval = −.86 to −.38).

*Does mean performance account for the association between variability and brain volume?*

The associations between variability and frontal lobe volumes remained significant after accounting for mean performance, F_{Change} (1, 38) = 4.48, $p = .041$ for the association between FW and block-to-block variability in CRT and F_{Change} (1, 38) = 15.04, $p < .001$ for the association between DLPFC and block-to-block variability in number–word memory accuracy. In line with these results and using the adjusted $p = .007$, we found no significant association between any regional brain volume and mean performance on any of the three cognitive tasks after statistically accounting for the influence of chronological age, all ps > .048.

Discussion

We observed that smaller prefrontal white matter volumes were associated with higher block-to-block variability in CRT, especially among older adults. Smaller volumes of DLPFC were associated with higher block-to-block variability in episodic memory (number–word pair) performance. This association was stronger in the group of younger adults. The observed associations between variability and brain volumes were not due to individual differences in mean performance.

Previous studies, focusing exclusively on trial-to-trial variability in RT tasks, have reported that greater variability in response speed is correlated with lower volume (Anstey et al., 2007; Walhovd and Fjell, 2007) and structural integrity (Bunce et al., 2007; Fjell et al., 2011; Moy et al., 2011; Tamnes et al., 2012) of white matter in several brain

---

[1] Contrary to the hypothesis, trial-to-trial variability in CRT was unrelated to regional volumes. We further probed this correlation by extracting trial-to-trial variability from one lexical-decision CRT task and one numerical magnitude CRT task that were administered at pretest. Importantly, these variability measures did not show unique associations with any regional brain volume after the influence of age was statistically controlled. Thus, the absence of volume–variability associations for these cognitive measures was not an artifact of a specific way of computing trial-to-trial variability.
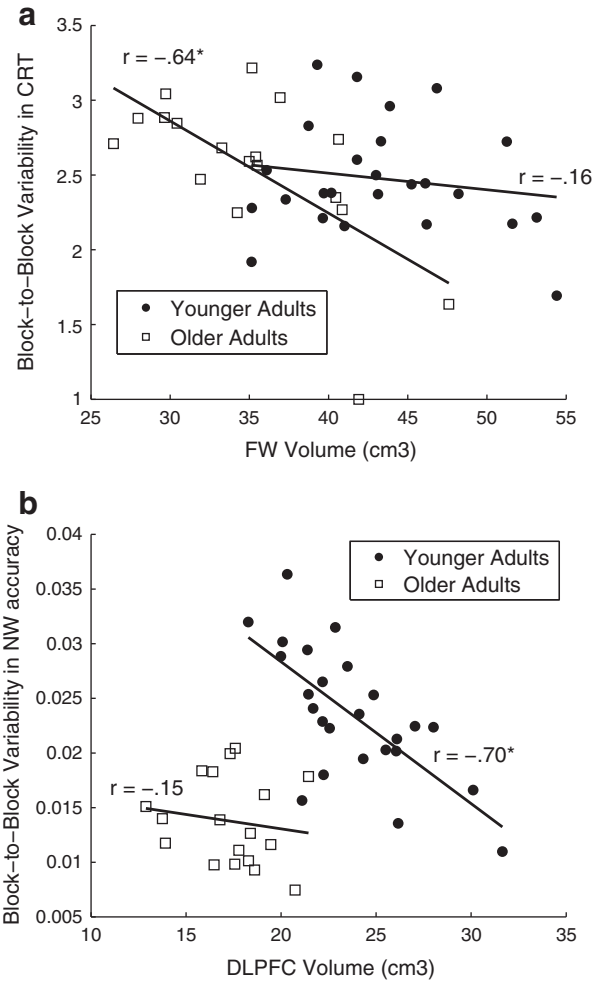
Fig. 1. Scatter plots of (A) the association between prefrontal white (FW) volume and block-to-block variability in the CRT task and (B) the association between dorsolateral prefrontal cortex (DLPFC) volume and block-to-block variability in number–word memory accuracy.

regions. In accord with these results, we observed an association between prefrontal white-matter volume and variability in response latency on a CRT task. In contrast to previous research, however, this association was not observed for trial-to-trial variability, but selectively for block-to-block variability in performance, which has not been assessed in previous studies. Note that, due to our variance partitioning approach, block-to-block and trial-to-trial variability are not confounded. That is, block-to-block variability in performance on the CRT task reflects only systematic changes in average performance from one block to another. In contrast, most previous studies on the neural correlates of variability in cognitive performance have confounded trial-to-trial and block-to-block variability. It is conceivable that fluctuations in frequencies below the trial-to-trial level have contributed to past estimates of trial-to-trial variability. Thus, the present results are important because they suggest that white matter volume is associated with variability in performance on cognitive tasks that draw heavily on speed of responding and that this association may originate from systematic variations in cognitive performance over blocks of trials, rather than from trial-to-trial fluctuations. Nevertheless, the failure to detect a significant association between trial-to-trial variability in performance and regional brain volume should not be over-interpreted, given the limited sample size in the present study.

The association between block-to-block variability in the CRT task and prefrontal white volume was greater in older adults than in their younger counterparts. In accord with a previous report on associations

between variability and white matter integrity (Fjell et al., 2011), a significant relationship emerged only in the older age group. One interpretation of this finding is that age-related decreases in prefrontal white matter volume results in increased variability in performance. There is ample evidence to suggest that inter-individual differences in brain and behavior in old age reflects, to a considerable degree, inter-individual differences in rates of age-related decline that are correlated across brain regions (Raz et al., 2005) and cognitive domains (de Frias et al., 2007; Ghisletta et al., 2012; Lindenberger and Ghisletta, 2009; Tucker-Drob, 2011). As a specific consequence of this general consideration, normal aging may amplify the covariation between white matter volume and cognitive performance. Consistent with this view is a recent report that individual differences in cortical thickness are larger and more predictive of individual differences in executive control in older adults relative to their younger counterparts (Burzynska et al., 2012). Similar considerations may apply to cognitive variability. In a cross-sectional study, regional brain volumes may reflect a multitude of influences related to genetic, prenatal and perinatal environments, maturation and development. In younger adults these factors should account for the bulk of the variance, whereas in older age groups, the differences in brain volumes should reflect, in addition, the influence of aging and age-related pathology. Thus, brain regions that show accelerated change in the older ages are more likely to show correlations with cognitive performance. For example, hippocampus and prefrontal white matter evidence accelerated age-related shrinkage (Raz et al., 2005), and indeed across multiple studies, hippocampal volume is associated with memory in the older but not younger adults (Van Petten, 2004). By the same logic, the correlations among between-person differences in prefrontal white matter volume and cognition are expected to increase with advanced age, which is what we observed in this study. In summary, we suggest that the observed age-related increase in the association between prefrontal white matter volume and block-to-block variability may be driven by developmental covariation between these two variables (see also Fjell et al., 2011). This interpretation is in line with empirical findings and conceptual considerations emphasizing the effects of correlated change on heterogeneity in old age (Hertzog, 1985; Hofer and Sliwinski, 2001; Lindenberger et al., in press; Lövdén and Lindenberger, 2005).

The observed association between block-to-block variability in episodic memory performance and DLPFC volume, and not any other gray matter volume, in healthy adulthood is a novel finding. This finding is consistent with the claims that short-term fluctuations in cognitive performance partially reflect lapses of attention, or transient failures in cognitive control (West et al., 2002; Williams et al., 2005), and that cognitive control relies on the dorsolateral prefrontal cortices (e.g. Miller and Cohen, 2001; Weissman et al., 2006). The observed link between DLPFC volumes and intra-individual variability in cognitive performance is also in accord with evidence from patients with DLPFC lesions (Stuss et al., 2003) and with results from work on animal models (Pouget et al., 2009). This finding, however, contradicts previous reports that fail to observe significant associations between trial-to-trial variability and gray matter differences as revealed by voxel-based morphometry (Moy et al., 2011; Ullén et al., 2008). Note also that recent findings indicate that variability on the millisecond time scale in motor tasks such as tapping may be related to cognitive performance in a manner that is partly independent from variability in RT (Holm et al., 2011). These, and other results (e.g., Ullén et al., 2012), suggest that bottom–up mechanisms (e.g., neural noise), may also induce associations among variability indices of cognitive performance and brain integrity. That is, our results should not be taken to suggest that fluctuation of cognitive control is the only source of cognitive variability, and of associations between variability and indices of brain integrity.

Of note is that this study differed from its predecessors in the manner of assessing variability. In the present report, we used accuracy of episodic memory performance across blocks of encoding and retrieving multiple items. Though such an index may be partly influenced by

trial-to-trial variability, block-to-block variability in accuracy tasks of this type may to a larger extent reflect influences such as search and selection of optimal strategies (Allaire and Marsiske, 2005; Li et al., 2004; Shing et al., 2012; Siegler, 1994). The distinct pattern of age differences in the association between block-to-block variability in episodic memory accuracy and DLPFC volume, with a significant association in the younger group only, can be interpreted to support this notion. That is, these results may indicate that experimentation with strategies influence observed block-to-block variability in episodic memory accuracy. The lower block-to-block variability in episodic memory accuracy for the older relative to the younger age group (see Fig. 1) further suggests that variability in this task may partly reflect adaptive rather than maladaptive antecedents of variability. In addition, follow up analyses indicate that, for the number–word task, mean and block-to-block variability was positively associated in the group of older adults ($r_{pretest} = .39$, $p = .109$; $r_{postest} = .60$, $p = .008$) but not among younger adults ($r_{pretest} = .03$, $p = .884$; $r_{postest} = -.12$, $p = .574$). These findings strongly suggest that, at least for older adults, variability in this task has an adaptive component. Individual differences in such adaptive sources of variability may co-exist with differences in other antecedents of variability that may reflect increased neural vulnerability, such as attentional lapses and neuronal noise. Depending on the mixture of such sources of variability, and between-subject differences in the various antecedents of variability, different associations between cognitive variability and brain volume may emerge. These results thus underscore the heterogeneity of the indices of variability in cognitive performance, and show that between-person differences in variability in a particular task may reflect a different mix of antecedents in different groups of participants (see also Allaire and Marsiske, 2005; Boker et al., 2009; MacDonald et al., 2009b; Ram et al., 2005). The negligible associations between day-to-day variability in cognitive performance and regional brain volumes further demonstrate the heterogeneity of measures of cognitive variability. This finding is consistent with the notion that such cognitive variability is governed more by non-cognitive factors (MacDonald et al., 2009b) such as day-to-day fluctuations in stress (Sliwinski et al., 2006) and motivation (Brose et al., 2010).

This study has several strengths, including the variance partitioning approach to uniquely estimate variability at different timescales (cf. Rabbitt et al., 2001), reliable and valid manual measurement of multiple regional brain volumes, and a wide range of different indices of variability in cognitive performance. However, a few limitations need to be acknowledged. Our sample was small and positively selected, reflecting recruitment based on convenience. We also note that we did not assess white matter volumes in the regions other than the frontal lobes. We therefore cannot determine whether the association between block-to-block variability in CRT performance with white matter volume is unique for frontal volume, or whether an association would also emerge for more posterior regions. Previous research would suggest that this association is not unique to frontal areas (Fjell et al., 2011; Ullén et al., 2008).

In summary, we have reported the novel results that block-to-block variability in episodic memory accuracy is related to DLPFC gray matter volume in younger adulthood, and that block-to-block variability (but not trial-to-trial variability) in CRT performance is related to prefrontal white volume in older adulthood. We conclude that between-person differences in multiple measures of variability in cognitive performance tap different constellations of behavioral and neural antecedents in distinct age groups.

Funding

Acknowledgments

References

Allaire, J.C., Marsiske, M., 2005. Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. Psychol. Aging 20, 390–401.

Anstey, K.J., Mack, H.A., Christensen, H., Li, S.C., Reglade-Meslin, C., Maller, J., Kumar, R., Dear, K., Easteal, S., Sachdev, P., 2007. Corpus callosum size, reaction time speed and variability in mild cognitive disorders and in a normative sample. Neuropsychologia 45, 1911–1920.

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. Neurosci. Biobehav. Rev. 30, 791–807.

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.

Bender, A.R., Raz, N., 2012. Age-related differences in memory and executive functions in healthy APOE ε4 carriers: the contribution of individual differences in prefrontal volumes and systolic blood pressure. Neuropsychologia 50, 704–714.

Boker, S.M., Molenaar, P.C., Nesselroade, J.R., 2009. Issues in intraindividual variability, individual differences in equilibria and dynamics over multiple time scales. Psychol. Aging 24, 858–862.

Brose, A., Schmiedek, F., Lovden, M., Molenaar, P.C.M., Lindenberger, U., 2010. Adult age differences in covariation of motivation and working memory performance, contrasting between-person and within-person findings. Res. Hum. Dev. 7, 61–78.

Bucur, B., Madden, D.J., Spaniol, J., Provenzale, J.M., Cabeza, R., White, L.E., Huettel, S.A., 2008. Age-related slowing of memory retrieval: contributions of perceptual speed and cerebral white matter integrity. Neurobiol. Aging 29, 1070–1079.

Bunce, D., Warr, P.B., Cochrane, T., 1993. Blocks in choice responding as a function of age and physical fitness. Psychol. Aging 8, 26–33.

Bunce, D., Anstey, K.J., Christensen, H., Dear, K., Wen, W., Sachdev, P., 2007. White matter hyperintensities and within-person variability in community-dwelling adults aged 60–64 years. Neuropsychologia 45, 2009–2015.

Burzynska, A.Z., Nagel, I.E., Preuschhof, C., Gluth, S., Bäckman, L., Li, S.-C., Lindenberger, U., Heekeren, H.R., 2012. Cortical thickness is linked to executive functioning in adulthood and aging. Hum. Brain Mapp. 33 (7), 1607–1620.

Craik, F.I.M., Salthouse, T.A. (Eds.), 2008. The Handbook of Aging and Cognition, 3rd ed. Psychology Press, New York, pp. 491–556.

de Frias, C.M., Lövden, M., Lindenberger, U., Nilsson, L.G., 2007. Revisiting the dedifferentiation hypothesis with longitudinal multi-cohort data. Intelligence 35, 381–392.

Espeseth, T., Greenwood, P.M., Reinvang, I., Fjell, A.M., Walhovd, K.B., Westlye, L.T., et al., 2006. Interactive effects of APOE and CHRNA4 on attention and white matter volume in healthy middle-aged and older adults. Cogn. Affect. Behav. Neurosci. 6, 31–43.

Fjell, A.M., Westlye, L.T., Amlien, I.K., Walhovd, K.B., 2011. Reduced white matter integrity is related to cognitive instability. J. Neurosci. 31, 18060–18072.

Folstein, M.F., Folstein, S.E., Mchugh, P.R., 1975. Mini-Mental State - Practical Method for Grading Cognitive State of Patients for Clinician. J. Psychiatr. Res. 12, 189–198.

Ghisletta, P., Rabbitt, P.M.A., Lunn, M., Lindenberger, U., 2012. Two thirds of the age-based changes in fluid and crystallized intelligence, perceptual speed, and memory in adulthood are shared. Intelligence 40 (3), 260–268.

Hertzog, C., 1985. An individual differences perspective: implications for cognitive research in gerontology. Res. Aging 7, 7–45.

Hofer, S.M., Sliwinski, M.J., 2001. Understanding ageing - An evaluation of research designs for assessing the interdependence of ageing-related changes. Gerontology 47, 341–352.

Holm, L., Ullen, F., Madison, G., 2011. Intelligence and temporal accuracy of behaviour: unique and shared associations with reaction time and motor timing. Exp. Brain Res. 214, 175–183.

Hoyer, W.J., Stawski, R.S., Wasylyshyn, C., Verhaeghen, P., 2004. Adult age and digit symbol substitution performance: a meta-analysis. Psychol. Aging 19, 211–214.

Hultsch, D.F., Strauss, E., Hunter, M., MacDonald, S.W.S., 2008. Intraindividual variability, cognition and aging, In: Craik, F.I.M., Salthouse, T.A. (Eds.), The Handbook of Aging and Cognition, 3rd ed. Psychology Press, New York, pp. 491–556.

Kennedy, K.M., Raz, N., 2009. Aging white matter and cognition: differential effects of regional variations in diffusion properties on memory, executive functions, and speed. Neuropsychologia 47, 916–927.

Kirchhoff, B.A., 2009. Individual Differences in Episodic Memory: The Role of Self-Initiated Encoding Strategies. Neuroscientist 15, 166–179.

Li, S.C., Lindenberger, U., Sikstrom, S., 2001. Aging cognition: from neuromodulation to representation. Trends Cogn. Sci. 5, 479–486.

Li, S.C., Huxhold, O., Schmiedek, F., 2004. Aging and attenuated processing robustness. Evidence from cognitive and sensorimotor functioning. Gerontology 50, 28–34.

Lindenberger, U., von Oertzen, T., 2006. Variability in cognitive aging: from taxonomy to theory. In: Bialystok, E., Craik, F.I.M. (Eds.), Lifespan Cognition: Mechanisms of Change. Oxford University Press, Oxford, pp. 297–314.

Lindenberger, U., Ghisletta, P., 2009. Cognitive and sensory declines in old age: gauging the evidence for a common cause. Psychol. Aging 24, 1–16.

Lindenberger, U., Mayr, U., Kliegl, R., 1993. Speed and intelligence in old age. Psychol. Aging 8, 207–220.

Lindenberger, U., Burzynska, A.Z., Nagel, I.E., in press. Heterogeneity in frontal-lobe aging (2nd ed.), in: Stuss, D.T., Knight, R.T. (Eds.), Principles of frontal lobe functions. Oxford University Press, New York.

Lövdén, M., Lindenberger, U., 2005. Development of intellectual abilities in old age: from age gradients to individuals. In: Wilhelm, O., Engle, R.W. (Eds.), Handbook of Understanding and Measuring Intelligence. Sage, Thousand Oaks, CA, pp. 203–221.

MacDonald, S.W.S., Nyberg, L., Backman, L., 2006. Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. Trends Neurosci. 29, 474–480.

MacDonald, S.W.S., Cervenka, S., Farde, L., Nyberg, L., Backman, L., 2009a. Extrastriatal dopamine D2 receptor binding modulates intraindividual variability in episodic recognition and executive functioning. Neuropsychologia 47, 2299–2304.

MacDonald, S.W.S., Li, S.C., Backman, L., 2009b. Neural underpinnings of within-person variability in cognitive functioning. Psychol. Aging 24, 792–808.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.

Moy, G., Millet, P., Haller, S., Baudois, S., de Bilbao, F., Weber, K., Lovblad, K., Lazeyras, F., Giannakopoulos, P., Delaloye, C., 2011. Magnetic resonance imaging determinants of intraindividual variability in the elderly: combined analysis of grey and white matter. Neuroscience 186, 88–93.

Nesselroade, J.R., 1991. The warp and woof of the developmental fabric. In: Downs, R.M., Liben, L.S., Palermo, D.S. (Eds.), Vision of aesthetics, the environment and development: the legacy of Joachim F. Wohlwill. Earlbaum, Hillsdale, NJ, pp. 213–240.

Pouget, P., Wattiez, N., Rivaud-Pechoux, S., Gaymard, B., 2009. A fragile balance: perturbation of GABA mediated circuit in prefrontal cortex generates high intraindividual performance variability. PLoS One 4, e5208.

Prado, J., Carp, J., Weissman, D.H., 2011. Variations of response time in a selective attention task are linked to variations of functional connectivity in the attentional network. Neuroimage 54, 541–549.

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. Q. J. Exp. Psychol. A 54, 981–1003.

Ram, N., Rabbitt, P., Stollery, B., Nesselroade, J.R., 2005. Cognitive performance inconsistency: intraindividual change and variability. Psychol. Aging 20, 623–633.

Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K.M., Williamson, A., Acker, J.D., 2004. Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. Neurobiol. Aging 25, 377–396.

Raz, N., Lindenberger, U., Rodrigue, K.M., Kennedy, K.M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., Acker, J.D., 2005. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. Cereb. Cortex 15, 1676–1689.

Schmiedek, F., Bauer, C., Lövdén, M., Brose, A., Lindenberger, U., 2010a. Cognitive enrichment in old age: Web-based training programs. GeroPsych 23, 59–68.

Schmiedek, F., Lövdén, M., Lindenberger, U., 2010b. Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO Study. Front. Aging Neurosci. 2.

Shing, Y.L., Schmiedek, F., Lövdén, M., Lindenberger, U., 2012. Memory updating training across 100 days in the COGITO study. Psychol. Aging 27, 451–461.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations — uses in assessing rater reliability. Psychol. Bull. 86, 420–428.

Siegler, R.S., 1994. Cognitive variability — a key to understanding cognitive development. Curr. Dir. Psychol. Sci. 3, 1–5.

Simmonds, D.J., Fotedar, S.G., Suskauer, S.J., Pekar, J.J., Denckla, M.B., Mostofsky, S.H., 2007. Functional brain correlates of response time variability in children. Neuropsychologia 45, 2147–2157.

Sliwinski, M.J., Smyth, J.M., Hofer, S.M., Stawski, R.S., 2006. Intraindividual coupling of daily stress and cognition. Psychol. Aging 21, 545–557.

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 126, 2363–2380.

Tamnes, C.K., Fjell, A.M., Westlye, L.T., Ostby, Y., Walhovd, K.B., 2012. Becoming consistent: developmental reductions in intraindividual variability in reaction time are related to white matter integrity. J. Neurosci. 32, 972–982.

Tucker-Drob, E.M., 2011. Global and domain-specific changes in cognition throughout adulthood. Dev. Psychol. 47, 331–343.

Ullén, F., Forsman, L., Blom, O., Karabanov, A., Madison, G., 2008. Intelligence and variability in a simple timing task share neural substrates in the prefrontal white matter. J. Neurosci. 28, 4238–4243.

Ullén, F., Söderlund, T., Kaaria, L., Madison, G., 2012. Bottom-up mechanisms are involved in the relation between accuracy in timing tasks and intelligence — further evidence using manipulations of state motivation. Intelligence 40, 100–106.

Van Petten, C., 2004. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. Neuropsychologia 42, 1394–1413.

Walhovd, K.B., Fjell, A.M., 2007. White matter volume predicts reaction time instability. Neuropsychologia 45, 2277–2284.

Wechsler, D., 1981. Manual for the Wechsler Adult Intelligence Scale — Revised. Psychological Corporation, New York.

Weissman, D.H., Roberts, K.C., Visscher, K.M., Woldorff, M.G., 2006. The neural bases of momentary lapses in attention. Nat. Neurosci. 9, 971–978.

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 Cogn. 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.