

Intraindividual Variability in Basic Reaction Time Predicts Middle-Aged and Older Pilots' Flight Simulator Performance

Quinn Kennedy,¹ Joy Taylor,^{1,2} Daniel Heraldez,¹ Art Noda,¹ Laura C. Lazzeroni,¹ and Jerome Yesavage,^{1,2}

¹Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, California.

²Department of Veterans Affairs Health Care System and Sierra-Pacific MIRECC, Palo Alto, California.

Objectives. Intraindividual variability (IIV) is negatively associated with cognitive test performance and is positively associated with age and some neurological disorders. We aimed to extend these findings to a real-world task, flight simulator performance. We hypothesized that IIV predicts poorer initial flight performance and increased rate of decline in performance among middle-aged and older pilots.

Method. Two-hundred and thirty-six pilots (40–69 years) completed annual assessments comprising a cognitive battery and two 75-min simulated flights in a flight simulator. Basic and complex IIV composite variables were created from measures of basic reaction time and shifting and divided attention tasks. Flight simulator performance was characterized by an overall summary score and scores on communication, emergencies, approach, and traffic avoidance components.

Results. Although basic IIV did not predict rate of decline in flight performance, it had a negative association with initial performance for most flight measures. After taking into account processing speed, basic IIV explained an additional 8%–12% of the negative age effect on initial flight performance.

Discussion. IIV plays an important role in real-world tasks and is another aspect of cognition that underlies age-related differences in cognitive performance.

Key Words: Cognitive aging—Intraindividual variability—Real-world performance.

THE purpose of this study was to investigate whether intraindividual variability (IIV), defined as moment-to-moment fluctuations in reaction times from one trial to another, predicts age-related changes in flight performance over time. A growing body of evidence indicates that IIV plays a key role in understanding age-related changes in fluid cognition and underlying neurological function (Bielak, Hultsch, Strauss, MacDonald, & Hunter, 2010b). First, there are consistent findings demonstrating greater IIV with increased age, even after controlling for age-related increases in mean reaction time (Bielak et al., 2010b; Bunce, MacDonald, & Hultsch, 2004; Deary & Der, 2005; MacDonald, Hultsch, & Dixon, 2003; Nesselroade & Salthouse, 2004; Salthouse, Nesselroade, & Berish, 2006; West, Murphy, Armilio, Craik, & Stuss, 2002). Age-related changes in IIV are seen in challenging cognitive tasks that involve ongoing attention such as episodic memory, inductive reasoning (Bielak, Hultsch, Strauss, Macdonald, & Hunter, 2010a; Bunce et al., 2004, 2007; Kelly, Uddin, Biswal, Castellanos, & Milham, 2008; MacDonald et al., 2003), as well as with mental fatigue (Bielak et al., 2010a). These findings are particularly relevant to this study, as flight control of an aircraft is a cognitively challenging task that requires sustained attention by the pilot.

Second, cross-sectional and longitudinal studies indicate that IIV is associated with cognitive ability. In a sample of healthy adults aged 60–71 years, those who had greater IIV on a simple reaction time task forgot a greater amount of information 1 week later than participants with lesser IIV (Papenberg et al., 2011). Additionally, age-related differences in performance on tasks including verbal/mental math task, computation span, letter series, word recall, story recall, and vocabulary disappeared or were substantially reduced after controlling for IIV (Deary & Der, 2005; MacDonald et al., 2003).

Longitudinal studies have found baseline IIV to predict cognitive decline and change in cognitive status in older adults several years later (Bielak et al., 2010a, 2010b; MacDonald et al., 2003). Bielak and colleagues (2010b) assessed whether baseline measures of IIV would predict cognitive decline 3 years later among a group of community-dwelling participants aged 64–92 years at baseline. Baseline measures of complex IIV (variability in reaction time (RT) for four choice one-back RT task and two-choice switch RT task) predicted cognitive decline in measures of recall, letter series, digit symbol, similarities, and vocabulary. Mixed modeling analyses indicated a significant negative covarying relationship between IIV and cognitive performance over

time. This result indicated that as IIV increased over time, cognitive performance correspondingly declined, with the strongest relationship with digit symbol. In another study by Bielak and colleagues (2010a) that utilized a subset of the sample mentioned earlier, Bielak and colleagues hypothesized that IIV would be a better predictor than mean RT of change in cognitive status and attrition over a 5-year time span. Although mean RT was a comparable predictor, the odds ratio for the IIV model was stronger. Baseline basic IIV successfully distinguished between participants who remained cognitively stable and those who showed decline over time. Similarly, for every 0.1 *SD* increase in basic IIV, the likelihood of attrition increased by 24%.

Finally, evidence points to IIV as a behavioral measure of neurophysiology and neurological functionality required for executive control processes (MacDonald, Cervenka, Farde, Nyberg, & Backman, 2009). MRI studies demonstrate that the frontal lobe gray and white matter are strongly associated with IIV (MacDonald et al., 2009). For example, white matter hyperintensities in the frontal lobe were positively associated with IIV in cognitively healthy participants aged 60–64 years (Bunce et al., 2007). Additionally, IIV from a simple reaction time task was significantly associated with corpus callosum size in participants in their early sixties; this association was stronger among participants with mild cognitive impairment (MCI) than healthy controls (Anstey et al., 2007). Regarding neurological function, IIV is negatively associated with efficiency of interhemispheric information processing and the correlation between task positive and task negative brain network activity (Anstey et al., 2007; Kelly et al., 2008) and is positively associated with dopamine dysregulation (MacDonald et al., 2009). *COMT* val carriers, who typically have decreased dopamine activity in the prefrontal cortex, show greater IIV on tasks that require cognitive stability than *met* carriers (MacDonald et al., 2009).

In summary, substantial evidence suggests that IIV plays an important and independent role in aging and cognitive function. One limitation of IIV findings is that the studies have focused on the association between IIV and performance on neuropsychological tests. It is unclear whether results extend to real-world tasks. The Stanford/VA Aviation lab has been annually assessing general aviators' performance in a flight simulator for the last 13 years, providing an ideal venue from which to test whether IIV results extend to a real world, complex task: flight performance. Our previous results have found that processing speed and executive function predict initial flight performance and rate of change in flight performance (Taylor, O'Hara, Mumenthaler, Rosen, & Yesavage, 2005; Yesavage et al., 2011). Thus, determining if IIV also affects flight performance is a logical next step in understanding in whom and under what circumstances flight performance may change.

In this study, we attempted to extend upon previous findings by investigating whether IIV predicts performance on a

cognitively complex, real-world task. Because we have also found processing speed and executive function to predict flight performance (Yesavage et al., 2011), measures of mean RT and executive function also were added to the model. The task assessed was flight simulator performance among middle-aged and older pilots who ranged in flight expertise. We hypothesized that IIV predicts initial level of flight performance and rate of decline in flight simulator performance even when mean RT and executive function are included

METHOD

Participants

This article reports findings on 236 pilots who were part of the ongoing longitudinal Stanford/VA Aviation Study approved by the Stanford University Institutional Review Board. Enrollment criteria were age between 40 and 69 years, current FAA medical certificate (Class III or higher), which entails an assessment of pilots' vision, hearing, and physical and mental health, and current flying activity between 300 and 15,000 hr of total flight time. All participants gave written informed consent to participate in the study, with the right to withdraw at any time. At entry, each participant was classified into one of three levels of aviation expertise depending on which Federal Aviation Administration (FAA) pilot proficiency ratings had been attained by study entry: (a) least expertise: VFR (rated for flying under visual flight rules only); (b) moderate expertise: IFR (also rated for instrument flight rules); and (c) most expertise: CFII and/or ATP (certified flight instructor of IFR students or rated for flying air-transport planes). As reported in our previous work (Taylor, Kennedy, Noda, & Yesavage, 2007), all of the VFR pilots were recreational pilots, although a small minority were employed in aviation-related jobs such as aircraft sales or mechanics. Within the IFR group, the majority were recreational pilots, whereas approximately one tenth were certified flight instructors, aviation analysts, or aviators during military service. Approximately, one half of the CFII/ATP participants were either air-transport pilots, CFIIIs, or their job duties included aircraft piloting.

Of the 277 pilots who had completed a test day by June 2011, 236 had complete data at the individual trial level for the cognitive measures of interest. Cogscreen AE (Kay, 1995) is programmed to automatically provide participant's mean reaction time and standard deviation across trials of a given measure. However, to attain the individual reaction times for each trial of a given measure required processing raw data files with a custom Perl program. Thus, although complete Cogscreen AE mean and standard deviation data are available for all 277 pilots, we were unable to retrieve the trial by trial individual reaction times for 41 pilots due to software malfunction. Pilots who were excluded from analyses due to missing cognitive data at the individual trial level did not significantly differ from study participants on

Table 1. Participants' Demographic and Flight Experience Information by Level of Expertise

	Pilot expertise level		
	Least (<i>n</i> = 62)	Moderate (<i>n</i> = 127)	High (<i>n</i> = 47)
Age in years, <i>M</i> (<i>SD</i>)	56.2 (7.3)	59.1 (6.4)	55.7 (6.5)
Years of education, <i>M</i> (<i>SD</i>)	16.5 (2.2)	17.2 (2.0)	17.2 (1.9)
Women, <i>n</i> (%)	5 (8.5)	21 (16.5)	4 (8.6)
White, non-Hispanic, %	57	126	43
Total log hours, <i>M</i> (<i>SD</i>)	1030 (1357)	1999 (2008)	5245 (2785)
Log hours in past month, mean (<i>SD</i>)	6.1 (7.4)	9.0 (10.8)	14.7 (15.6)
Family history of dementia, %: "no/yes/not sure"	79/21/0	60/32/9	79/21/0

Note. Log hours are the flight hours pilots document in their log books; that is, log hours are a measure of expertise.

flight expertise classification, total flight hours, years of education, or age (p 's > .39). Of the remaining 236 pilots, 30 were women, 62 had VFR ratings, 127 had IFR/CFI ratings, and 47 had ATP/CFII ratings. These participants had an average of 3.56 (2.47) annual flight simulator assessments (range = 1–12 assessments). Table 1 provides demographic and flight experience characteristics of the sample.

Equipment

Pilots "flew" in a Frasca 141 flight simulator (Urbana, IL). Motion, vibration, and sound elements were not incorporated into this simulator protocol. The simulator was linked to a computer specialized for graphics (Dell Precision Workstation and custom C++ OpenGL Linux software) that generated a "through-the-window" visual environment and continuously collected data concerning the aircraft's position and communication frequencies. The simulator is located in a quiet, darkened room kept at a comfortable temperature with the cockpit independently lit from the projector display. The display is projected on a screen 15' in front of the pilot. The simulation occurred during normal working hours from 0900 to 1600 at the pilot's preference. Previous work in our lab indicates that the flight simulator has validity as it distinguishes performance between novice and expert aviators and between younger and older aviators (Taylor et al., 2005, 2007).

Measures

Flight simulator performance.—The scoring system of the flight simulator–computer system produces 23 variables that measure deviations from ideal positions or assigned values (e.g., altitude in feet, heading in degrees, airspeed in knots), or reaction time in seconds (Yesavage, Taylor, Mumenthaler, Noda, & O'Hara, 1999). Because these individual variables have different units of measurement, the raw scores for each variable were converted to z scores using the baseline visit mean and standard deviation of 141 participants enrolled during 1996–2001 (scores on the morning and afternoon flights were averaged). The z scores on the individual measures were aggregated on the basis of previous principal component analyses into four component measures (Yesavage et al., 1999, 2002): (a) accuracy

of executing the air traffic control (ATC) communications regarding the heading, altitude, radio frequency, and transponder code; (b) traffic avoidance; (c) scanning cockpit instruments to detect engine emergencies; and (d) executing a visual approach to landing. A flight summary score, the average of the above four component measures, was used as the primary performance measure. Thus, one global and four component measures of flight performance were assessed.

IIV measures.—Subtests of Cogscreen-AE (Kay, 1995), a computerized battery of cognitive tests specifically geared to aviators were used.

Pathfinder.—A sequencing and visual scanning task. The participant uses a light pen to (a) sequentially connect numbers (Pathfinder Number), (b) connect letters in alphabetic order (Pathfinder Letters), and (c) sequence an alternating set of numbers and letters (Pathfinder combined).

Shifting attention.—The Shifting Attention Test (SAT) is designed to measure the ability to maintain attentional set and shift between sets. Each of the randomly generated probe stimuli consist of a square with a surrounding border that is either purple or yellow; inside the square is a yellow or purple arrow that is pointing to the left or to the right. The four possible response choices remain constant across trials. The participant's task is to touch the response box that matches the probe according to the current "rule." The rule can be as follows: response based on border color, arrow color, or arrow direction. Response times during three of the five SAT subtests were used in measuring IIV: (a) arrow direction, the ability to select correct box based on direction of the arrow, (2) arrow color, the ability to select correct box based on color of the arrow, and (3) instruction, the ability to correctly apply an instruction that cued the correct rule prior to each stimulus presentation.

Divided attention test indicator alone task.—A visual-motor tracking task in which the participant uses a light pen to center a vertically drifting cursor. When the cursor is in the upper or lower sections of the area, the participant must touch a response box to return the cursor to the center

section. The median amount of time the cursor spends outside the center section before the participant begins to move the cursor is recorded.

Symbol digit coding task.—A touch screen analogue of the Symbol Digit Modalities Test. A set of symbol–digit pairings is shown continuously on the screen during the task. Probe symbols are presented one at a time, and the participant is supposed to point to the digit below that corresponds to the symbol. The participant is given 90 s to complete as many items as possible.

Executive function measure.—The Discovery subtest of the Shifting Attention Test was used to measure executive function. The Discovery subtest was presented after the other four SAT subtests. As in the Wisconsin Card Sorting Test, participants use trial and error to discover which stimulus dimension (such as arrow color) is currently relevant and then respond according to that rule until feedback indicates that it is no longer relevant. Three types of performance were measured: (a) number of completed rule sets, (b) number of failures to maintain set, and (c) the percentage of correct responses. These three performance measures were standardized and averaged into a composite measure of executive function.

Processing speed measure.—Processing speed was a composite measure of speeded performance during 11 visual scanning and perceptual comparison tasks found in Cogscreen-AE (Kay, 1995). Performance on all of these tasks is measured as response “throughput,” which is the number of correct responses made per minute. Eight of the 11 tasks were throughput components of the pathfinder, shifting attention, divided attention, and symbol digit tasks described earlier. The other tasks were visual sequence comparison, matching to sample, and manikin. Full descriptions of these other tasks are available online (<http://www.cogscreen.com/>) and in the CogScreen-AE manual (Kay, 1995).

Data Preparation

Five steps were taken to create the IIV measures. Step 1 entailed a principal component analysis (PCA) of participants’ performance on the cognitive variables at their initial visit so as to create composite IIV measures. The PCA with Spearman’s correlations found two factors. The first factor was comprised of Pathfinder Number, Pathfinder Letter, and Pathfinder Combined (factor loadings ranged from .704 to .892; variance explained: 2.510). We refer to this factor as basic IIV. The second factor was comprised of shifting attention arrow color throughput, shifting attention arrow direction throughput, shifting attention instruction throughput, divided attention indicator alone speed, and symbol digit coding throughput (factor loadings ranged from .529 to .765, variance explained: 2.358). We refer to this factor as complex IIV.

For steps 2–5, we followed similar data preparation procedures as Bielak and colleagues (2010a, 2010b). Step 2 involved removing the high and low outliers in reaction time from each cognitive variable for each participant. High outliers were defined as individual reaction times that were greater than 3 *SD* more than the person’s mean reaction time for that particular test. Low outliers were defined as individual reaction times less than 150 ms. After the outliers were removed, mean RT and within-person individual standard deviations (ISDs) were recalculated for each participant and cognitive variable. ISDs of all variables were normally distributed.

Step 3 entailed removing the effect of mean reaction time from the ISDs because mean RT is positively associated with variability in reaction time and age is associated with slower reaction times (Anstey et al., 2007; Hultsch, MacDonald, & Dixon, 2002). For each variable, the ISDs were regressed on mean reaction time, and the residuals were saved. Step 3 also ensured that the IIV and mean RT would be independent predictors in the model used for hypothesis testing. Step 4 involved standardizing the residuals into *z* scores. For each participant, the standardized ISD residuals from each of the nine cognitive variables were collected. Finally, step 5 consisted of creating basic IIV and complex IIV composite scores of the standardized residuals. Composite scores were created by averaging together the standardized ISD residuals from the cognitive variables for each factor. Thus, there was one composite score of standardized ISD residuals based on the pathfinder reaction times comprising the basic IIV factor and another composite score of standardized ISD residuals based on reaction times from measures comprising the complex IIV factor.

No significant correlations were found between the processing speed measure and the IIV variables (p 's > .19). The executive function measure was marginally correlated with basic IIV (Spearman’s $r = -.13$, $p = .051$, $n = 220$). Age was slightly correlated with the basic IIV (Spearman’s $r = .16$, $p = .018$, $n = 235$) and negatively correlated with the processing speed measure (Spearman’s $r = -.39$, $p < .0001$, $n = 235$) and the executive function measure (Spearman’s $r = -.18$, $p = .007$, $n = 220$). The two IIV variables were moderately correlated with each other (Spearman’s $r = .35$, $p < .0001$, $n = 236$).

Statistical Analyses

Preliminary analyses indicated that the complex IIV factor was not correlated with any outcome variables. Because it also was significantly correlated with basic IIV, it was dropped from analyses used for hypothesis testing. To test the hypothesis regarding basic IIV as predictor of initial level of flight performance and rate decline in flight performance, we conducted mixed modeling assuming a linear trend of performance over time (i.e., age), in which the full model had the following predictors: flight expertise

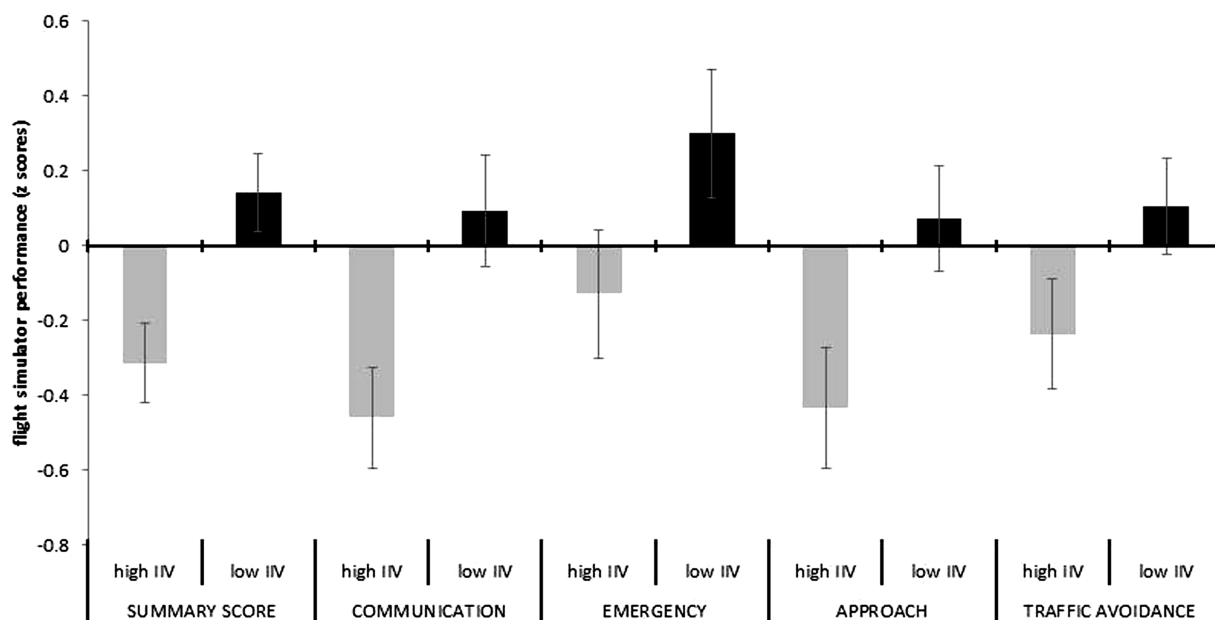


Figure 1. Participants with the greatest intraindividual variability (IIV) perform worse on initial flight simulator measures than those with the least IIV (greatest 10% and least 10% n 's = 24).

and initial measures of IIV, processing speed, and executive function (PROC MIXED procedure in SAS software version 9.1.3 [Cary, NC]). In the model, age is represented by the intercept for both initial and rate of decline in flight performance. We also conducted the analysis allowing for a nonlinear trend (by adding an age \times age term in the model). However, allowing for a nonlinear trend did not improve the fit of the model; therefore, the model with a linear trend was used in hypothesis testing.

Procedure

Participants had one 45-min practice flight in the simulator to experience the simulator's flight and landing characteristics. Additionally, participants completed five 75-min practice flights to gain familiarity with the flight scenario used throughout the study. Participants typically completed two practice flights a day during a 1- to 3-week period, after which they had a 3-week break before returning for the test day. During the test day, the participant flew a 75-min flight in the morning and a 75-min flight in the afternoon. Each flight was followed by a 40- to 60-min battery of cognitive tests, including CogScreen-AE (Kay, 1995), a computer-administered battery of 13 tests designed to assess perceptual and cognitive abilities relevant to aircraft piloting. The entire test day lasted approximately 6 hr, including a 30- to 50-min lunch break. Each flight began with the ATC's takeoff clearance. The first ATC message was presented 3 min later, after participants had lifted off the runway and climbed to 1200 ft. (365.76 m). During the flight, pilots heard 16 ATC messages, presented at the rate of one message every 3 min, directing the pilot to fly a new heading,

a new altitude, dial in a new radio frequency, and in 50% of the legs, dial in a new transponder code. Participants were instructed to read back the ATC messages and execute them in order and according to FAA standards. To further increase workload, pilots were confronted with randomly presented emergency situations: engine malfunctions (carburetor icing, drop of engine oil pressure; 8 of 16 legs) and/or suddenly approaching air traffic (10 of 16 legs). Pilots were to report engine malfunctions immediately and to avoid air traffic by veering quickly yet safely in the direction diagonal to the path of the oncoming plane. Pilots flew in severe turbulence throughout the flight and also encountered a 15-knot crosswind during approach and landing. Multiple versions of this flight scenario were presented to reduce learning of specific maneuvers and ATC messages.

RESULTS

Our hypothesis was partially supported. Although basic IIV did not predict rate of decline in flight simulator performance, it was a significant predictor for initial flight simulator performance on all measures except for approach, which trended toward significance. Figure 1 illustrates these results by comparing initial flight simulator performance for participants who were above the 90th percentile for basic IIV (high variability in reaction time) and participants who were below the 10th percentile for basic IIV (low variability in reaction time). In addition, consistent with our previous results, age, expertise, processing speed, and executive function also predicted initial performance (Kennedy, Taylor, Reade, & Yesavage, 2010; Taylor et al., 2007; Yesavage et al., 2011). Table 2 presents results from

Table 2. Mixed Effects Growth Curve Analysis of Longitudinal Flight Performance^a

	Summary score	Communication	Traffic avoidance	Emergency	Approach
	Parameter estimate (SE)				
Initial performance ^b (<i>I</i> _{<i>i</i>})					
Intercept (mean, η_1)	0.011 (0.026) (<i>p</i> = .6791)	-0.113 (0.041) (<i>p</i> = .0058)	0.155 (0.0381) (<i>p</i> < .0001)	0.0325 (0.054) (<i>p</i> = .5452)	-0.039 (0.035) (<i>p</i> = .2642)
Expertise (β_{11})	0.171 (0.034) (<i>p</i> < .0001)	0.259 (0.052) (<i>p</i> < .0001)	0.132 (0.048) (<i>p</i> = .0070)	0.120 (0.069) (<i>p</i> = .0847)	0.188 (0.044) (<i>p</i> < .0001)
IIV (β_{12})	-0.194 (0.035) (<i>p</i> < .0001)	-0.266 (0.053) (<i>p</i> < .0001)	-0.190 (0.050) (<i>p</i> = .0002)	-0.219 (0.070) (<i>p</i> = .0021)	-0.080 (0.045) (<i>p</i> = .078)
Processing speed (β_{13}) ^c	0.345 (.041) (<i>p</i> < .0001)	0.435 (0.062) (<i>p</i> < .0001)	0.265 (0.0589) (<i>p</i> < .0001)	0.504 (0.083) (<i>p</i> < .0001)	0.175 (0.055) (<i>p</i> = .0016)
Executive function (β_{14})	0.062 (.027) (<i>p</i> = .0212)	0.095 (0.041) (<i>p</i> < .0208)	0.033 (0.038) (<i>p</i> = .3857)	0.0853 (.054) (<i>p</i> = .1156)	0.019 (0.035) (<i>p</i> = .5952)
Change in performance over age ^d (<i>S</i>)					
Intercept (mean, η_5)	-0.011 (0.003) (<i>p</i> = .0003)	-0.021 (0.004) (<i>p</i> < .0001)	0.003 (0.005) (<i>p</i> = .5719)	0.001 (0.006) (<i>p</i> = .8554)	-0.028 (0.005) (<i>p</i> < .0001)
Expertise β_{51}	-0.001 (0.004) (<i>p</i> = .8753)	-0.005 (0.006) (<i>p</i> = .4015)	0.002 (0.007) (<i>p</i> = .7427)	-0.004 (.008) (<i>p</i> = .6517)	0.006 (0.006) (<i>p</i> = .3032)
IIV (β_{52})	0.006 (0.004) (<i>p</i> = .158)	0.010 (0.006) (<i>p</i> = .0900)	0.008 (.006) (<i>p</i> = .1867)	0.004 (0.008) (<i>p</i> = .6360)	-0.001 (0.006) (<i>p</i> = .8452)
Processing speed (β_{53})	0.003 (.005) (<i>p</i> = .4846)	0.000 (0.007) (<i>p</i> = .9958)	-0.0003 (0.008) (<i>p</i> = .9655)	0.006 (0.009) (<i>p</i> = .5113)	0.011 (.007) (<i>p</i> = .1435)
Executive function(β_{54})	0.000 (.003) (<i>p</i> = .9491)	0.005 (0.005) (<i>p</i> = .2510)	0.008 (0.005) (<i>p</i> = .0996)	-0.012 (0.006) (<i>p</i> = .0600)	-0.000 (0.005) (<i>p</i> = .9779)

Notes. ^aThe model for the outcome at a given age was $Y_{it} = I_i + S \times (\text{age centered}) + e_{it}$, in which the outcome *Y* for individual *i* at age *t* is a function of random initial status *I_i*. The residual *e_{it}* is assumed to be normally distributed.

^bThe model for initial performance was $I_i = \eta_1 + \beta_{11} \times (\text{expertise centered}) + \beta_{12} \times (\text{basic IIV } z \text{ scored}) + \beta_{13} \times (\text{processing speed centered}) + \beta_{14} \times (\text{executive function centered}) \zeta_{1i}$.

^cProcessing speed was measured as a throughput measure: positive coefficients indicate better flight performance.

^dThe model for the rate of decline (slope) in performance was $S = \eta_5 + \beta_{51} \times (\text{expertise centered}) + \beta_{52} \times (\text{basic IIV } z \text{ scored}) + \beta_{53} \times (\text{processing speed centered}) + \beta_{54} \times (\text{executive function centered})$. The random effect residual ζ_{1i} is assumed to be normally distributed.

the full model. The upper section of Table 2 shows the initial level of flight simulator performance and predictors of initial performance. The lower section of Table 2 shows the relationship between age-related decline in flight simulator performance and predictors of rate of decline in performance. As in our previous work (Taylor et al., 2007), we found significant rate of decline in flight simulator performance for the overall summary score, communications, and the approach variables.

We next conducted exploratory analyses to determine how much of the age-related variability in initial flight performance can be explained by IIV. We followed the procedure outlined in Taylor and colleagues (2005), in which four hierarchical linear models (HLM) were conducted on each of the flight performance measures from participants' initial visit. In our exploratory analyses, the predictors and their order in each HLM were as follows: model 1, age; model 2, processing speed and age; model 3, basic IIV and age; model 4, processing speed, basic IIV, and age. A ratio of the type I sums of squares (SS) for age were then calculated for models 2, 3, and 4; for example, for model 2, we calculated the percentage decrease in age-related variance (ARV) as

$$\% \text{ decrease in ARV} = 100 - \left[\left(\frac{\text{Type I SS Age}_{(\text{from model 2})}}{\text{Type I SS Age}_{(\text{from model 1})}} \right) \times 100 \right]$$

When basic IIV was added to the regression model with age (model 2), the ARV was reduced by between 15% and 22% across baseline flight performance measures. Results from model 3 indicated that mean RT accounted for 58%–88% of the variability in age. Together, IIV and mean RT reduced the amount of ARV from 70% to 82% across flight performance measures. Although IIV and mean RT together accounted for 96% of the ARV in emergencies, it should be noted that this model was not significant. Table 3 describes these results.

DISCUSSION

Results from a sample of middle-aged and older pilots suggest that IIV provides additional information regarding cognitive processing speed beyond mean RT. For almost all initial flight performance measures, IIV on basic reaction time tasks was a significant predictor, even after measures of mean RT and executive function were included in the model. Additionally, basic IIV explained between 15% and 22% of the ARV in initial flight simulator performance. Although mean RT explained a greater proportion of age-related variability in flight performance (between 58% and 70%), it is important to note that adding basic IIV to the model consistently explained an additional 11%–12% of the ARV. These results demonstrate that findings of an

Table 3. Age-Related Differences in Initial Flight Simulator Performance are Reduced by Basic **Intraindividual Variability** (IIV) and Mean RT

Variables	Model information	Model 1		Model 2		Model 3		Model 4	
		Age	Mean RT	Age	Basic IIV	Age	Mean RT	Basic IIV	Age
Summary score									
	Type I SS	11.01	17.11	3.32	4.73	9.13	17.11	5.27	2.06
	Incremental R^2	0.16	0.25	0.05	0.07	0.14	0.26	0.08	0.03
	% decrease in ARV			69.84		17.05			81.31
	Model p value	$p < .0001$		$p < .0001$		$p < .0001$		$p = .001$	
Communication									
	Type I SS	19.97	32.06	5.84	8.59	16.56	32.06	9.59	3.58
	Incremental R^2	0.15	0.24	0.04	0.07	0.13	0.24	0.07	0.03
	% decrease in ARV			70.79		17.07			82.08
	Model p value	$p < .0001$		$p = .0002$		$p < .0001$		$p = .0022$	
Traffic avoidance									
	Type I SS	9.16	12.57	3.1	3.6	7.67	12.57	4.01	2.03
	Incremental R^2	0.10	0.13	0.03	0.04	0.08	0.13	0.04	0.02
	% decrease in ARV			66.15		16.28			77.83
	Model p value	$p < .0001$		$p = .0027$		$p < .0001$		$p = .0131$	
Emergency									
	Type I SS	4.44	12.85	0.55	2.99	3.47	12.85	3.37	0.18
	Incremental R^2	0.02	0.07	0.00	0.02	0.02	0.07	0.02	0.00
	% decrease in ARV			87.62		21.67			96.03
	Model p value	$p = .0207$		$p = .4031$		$p = .0398$		$p = .6340$	
Approach									
	Type I SS	13.46	14.11	5.62	4.56	11.45	14.11	5.04	4.00
	Incremental R^2	0.11	0.12	0.05	0.04	0.10	0.12	0.04	0.03
	% decrease in ARV			58.26		14.98			70.27
	Model p value	$p < .0001$		$p = .0004$		$p < .0001$		$p = .0022$	

Note. ARV = age-related variance. $N = 235$.

IIV effect on computerized cognitive tests extend to a real-world task among a group of middle-aged and older adults with specialized training.

Greater variability in reaction time had an adverse impact on the ability of the pilot to maintain control of the aircraft. Many aviation tasks, such as those assessed in our study, require sustained monitoring of the environment out the window and the relevant instruments while simultaneously fine tuning aircraft control inputs to maintain course. This need for ongoing attention may be heightened during ATC communications, in which the pilot has the additional cognitive burden of remembering and executing the commands. Similarly, the pilot has to react in a timely fashion when oncoming traffic is detected. Another real-world task that may be affected by IIV is driving, particularly among middle-aged and older adults, as driving entails similar cognitive demands as flight control.

Unlike other studies, complex IIV was not associated with any flight performance measure. One possible reason is that our composite measure of complex IIV encompassed multiple tasks that entail both shifting attention and divided attention. In studies that found an effect of complex IIV, complex IIV was measured more parsimoniously, that is, a composite of the four choice reaction time one-back task and the two-choice switch reaction time task (Bielak et al., 2010a, 2010b).

Also inconsistent with previous findings, initial IIV did not predict rate of decline in flight performance (Bielak et al., 2010a, 2010b; MacDonald et al., 2003). Previous studies spanned at least 5 years. In contrast, the majority of our participants had 3 years of data. A longer time span may be necessary to detect IIV effects on longitudinal performance on real-world tasks. In our previous work, executive function and processing speed predicted who would show steeper decline in performance (Yesavage et al., 2011). However, those results are based on a much larger sample size than in this study (due to software malfunction, we were unable to retrieve individual trial by trial reaction times for 42 potential participants).

In summary, the results demonstrate that IIV is another aspect of cognition that underlies age-related differences in cognitively demanding tasks independently of mean reaction time and executive function. IIV had a negative effect on multiple aspects of flight control, which is a crucial skill for safe flying. To more deeply understand the underlying neurological basis of these results, we plan to compare greater IIV and lesser IIV pilots on neurophysiological and neurological functionality measures (fMRI, eyetracking). Although flying is a specialized skill, the results suggest that IIV may also affect driving performance, as driving requires many of the same attentional demands as flying. As an increasing proportion of drivers

are older adults, future studies should investigate the effect of IIV on driving.

FUNDING

This research is supported in part by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, the Department of Veterans Affairs Sierra-Pacific Mental Illness Research, Education, and Clinical Center (MIRECC), the War Related Illness and Injury Study Center (WRIISC), and by Grant Number R37 AG 12713 from the National Institute on Aging. These sponsors solely provided financial support or facilities to conduct the study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

ACKNOWLEDGMENTS

We thank the study's paid research assistants, including Katy Castile of Stanford University, for recruiting and testing participants. We also thank the aviator study participants for their donation of time and for being inspirational role models of intellectual exploration.

CORRESPONDENCE

Correspondence should be addressed to Quinn Kennedy, PhD, Stanford/VA Aviation Lab, VA Palo Alto Health Care System (151Y), 3801 Miranda Avenue, Palo Alto, CA 94304-1207. E-mail: quinnk@stanford.edu.

REFERENCES

- Anstey, K. J., Mack, H. A., Christensen, H., Li, S. C., Reglade-Meslin, C., Maller, J., ... Sachdev, P. (2007). Corpus callosum size, reaction time speed and variability in mild cognitive disorders and in a normative sample. *Neuropsychologia*, *45*, 1911–1920. doi:10.1016/j.neuropsychologia.2006.11.020
- Bielak, A. A., Hulstsch, D. F., Strauss, E., MacDonald, S. W., & Hunter, M. A. (2010a). Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. *Neuropsychology*. doi:10.1037/a0019802
- Bielak, A. A., Hulstsch, D. F., Strauss, E., MacDonald, S. W., & Hunter, M. A. (2010b). Intraindividual variability is related to cognitive change in older adults: Evidence for within-person coupling. *Psychology and Aging*, *25*, 575–586. doi:10.1037/a0019503
- 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. doi:10.1016/j.neuropsychologia.2007.02.006
- Bunce, D., MacDonald, S. W., & Hulstsch, D. F. (2004). Inconsistency in serial choice decision and motor reaction times dissociate in younger and older adults. *Brain and Cognition*, *56*, 320–327. doi:10.1016/j.bandc.2004.08.006
- 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. doi:10.1080/13825580590969235
- Hulstsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology. Series B, Psychological Sciences and Social Sciences*, *57*, P101–P115.
- Kay, G. G. (1995). *CogScreen Aeromedical edition professional manual*. Odessa, FL: Psychological Assessment Resources.
- Kelly, A. M., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2008). Competition between functional brain networks mediates behavioral variability. *Neuroimage*, *39*, 527–537. doi:10.1016/j.neuroimage.2007.08.008
- Kennedy, Q., Taylor, J. L., Reade, G., & Yesavage, J. A. (2010). Age and expertise effects in aviation decision making and flight control in a flight simulator. *Aviation, Space, and Environmental Medicine*, *81*, 489–497.
- MacDonald, S. W., Cervenka, S., Farde, L., Nyberg, L., & Backman, L. (2009). Extrastriatal dopamine D2 receptor binding modulates intraindividual variability in episodic recognition and executive functioning. *Neuropsychologia*, *47*, 2299–2304. doi:10.1016/j.neuropsychologia.2009.01.016
- MacDonald, S. W., Hulstsch, 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. doi:10.1037/0882-7974.18.3.510
- Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *59*, P49–P55.
- Papenberg, G., Backman, L., Chicherio, C., Nagel, I. E., Hauke, R., Lindenberger, U., & Li, S. (2011). Higher intraindividual variability is associated with more forgetting and dedifferentiated memory functions in old age. *Neuropsychologia*, *49*, 1879–1888.
- Salthouse, T. A., Nesselroade, J. R., & Berish, D. E. (2006). Short-term variability in cognitive performance and the calibration of longitudinal change. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *61*, P144–P151. doi:10.1093/geronb/61/3/P144 [pii]
- Taylor, J. L., Kennedy, Q., Noda, A., & Yesavage, J. A. (2007). Pilot age and expertise predict flight simulator performance: a 3-year longitudinal study. *Neurology*, *68*, 648–654. doi:10.1212/01.wnl.0000255943.10045.c0
- Taylor, J. L., O'Hara, R., Mumenthaler, M. S., Rosen, A. C., & Yesavage, J. A. (2005). Cognitive ability, expertise, and age differences in following air-traffic control instructions. *Psychology and Aging*, *20*, 117–133.
- 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. doi:S0278262601915076 [pii]
- Yesavage, J. A., Jo, B., Adamson, M. M., Kennedy, Q., Noda, A., Hernandez, B., ... Taylor, J. L. (2011). Initial cognitive performance predicts longitudinal aviator performance. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *66*, 444–453. doi:10.1093/geronb/gbr031
- Yesavage, J. A., Mumenthaler, M. S., Taylor, J. L., Friedman, L., O'Hara, R., Sheikh, J., ... Whitehouse, P. J. (2002). Donepezil and flight simulator performance: effects on retention of complex skills. *Neurology*, *59*, 123–125.
- Yesavage, J. A., Taylor, J. L., Mumenthaler, M. S., Noda, A., & O'Hara, R. (1999). Relationship of age and simulated flight performance. *Journal of the American Geriatrics Society*, *47*, 819–823.