

Research Article

When I'm 64: Age-Related Variability in Over 40,000 Online Cognitive Test Takers

Annalise A. LaPlume, PhD,^{1,✉} Nicole D. Anderson, PhD,^{1,2,✉} Larissa McKetton, PhD,¹ Brian Levine, PhD,^{1,3} and Angela K. Troyer, PhD^{4,5,*}

¹Rotman Research Institute, Baycrest Health Sciences (fully affiliated with the University of Toronto), Ontario, Canada. ²Departments of Psychology & Psychiatry, University of Toronto, Ontario, Canada. ³Departments of Psychology & Medicine (Neurology), University of Toronto, Ontario, Canada. ⁴Department of Psychology, University of Toronto, Ontario, Canada. ⁵Neuropsychology and Cognitive Health Program, Baycrest Health Sciences, Toronto, Ontario, Canada.

*Address correspondence to: Angela K. Troyer, Neuropsychology and Cognitive Health Program, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada. E-mail: atroyer@baycrest.org

Received: February 16, 2021; Editorial Decision Date: July 16, 2021

Decision Editor: Alyssa Gamaldo, PhD

Abstract

Objectives: Age-related differences in cognition are typically assessed by comparing groups of older to younger participants, but little is known about the continuous trajectory of cognitive changes across age, or when a shift to older adulthood occurs. We examined the pattern of mean age differences and variability on episodic memory and executive function measures over the adult life span, in a more fine-grained way than past group or life-span comparisons.

Method: We used a sample of over 40,000 people aged 18–90 who completed psychometrically validated online tests measuring episodic memory and executive functions (the Cogniciti Brain Health Assessment).

Results: Cognitive performance declined gradually over adulthood, and rapidly later in life on spatial working memory, processing speed, facilitation (but not interference), associative recognition, and set shifting. Both polynomial and segmented regression fit the data well, indicating a nonlinear pattern. Segmented regression revealed a shift from gradual to rapid decline that occurred in the early 60s. Variability between people (interindividual variability or diversity) and variability within a person across tasks (intraindividual variability or dispersion) also increased gradually until the 60s, and rapidly after. Confirmatory factor analysis revealed a single general factor (of variance shared between tasks) offered a good fit for performance across tasks.

Discussion: Life-span cognitive performance shows a nonlinear pattern, with gradual decline over early and mid-adulthood, followed by a transition in the 60s to notably accelerated, but more variable, decline. Some people show less decline than others, and some cognitive abilities show less within-person decline than others.

Keywords: Aging, Cognition, Episodic memory, Executive function, Life span

A body of evidence shows that while crystallized cognitive ability is relatively intact into later life (Craik & Bialystok, 2006; Hartshorne & Germine, 2015), older adults perform worse than younger adults on measures of fluid cognition, including working memory (Bopp & Verhaeghen, 2005), interference control (Rey-Mermet & Gade, 2018), and associative recognition memory (Old &

Naveh-Benjamin, 2008). However, the typical approach of comparing younger and older adult groups offers little insight into how and when decline occurs across the adult life span. Some researchers claim that decline occurs monotonically over midlife, whereas others report that cognition remains relatively stable in midlife and declines later in life (Christensen et al., 1994). There is also

a lack of consensus on the timing of accelerated decline in later life, ranging from an onset in the fourth decade to the seventh decade (Christensen et al., 1994; Morse, 1993). Finally, what constitutes older adulthood varies across studies, with the start of older adulthood ranging from 50 to 70.

Life-span studies show a steady worsening of performance per decade on various cognitive measures, including working memory (Borella et al., 2008; Chiappe et al., 2000; Myerson et al., 2003; Park et al., 2002), processing speed (Salthouse et al., 2000); interference control (Bugg et al., 2007; Troyer et al., 2006; Uttl & Graf, 1997), associative recognition memory (Bender et al., 2010), and set shifting (Periáñez et al., 2007; Salthouse et al., 2000; Tombaugh, 2004). Polynomial models often account for effects of age on cognition better than linear models, indicating that decline accelerates later in life (Borella et al., 2008; Myerson et al., 2003; Salthouse & Meinz, 1995; Uttl & Graf, 1997).

Challenges in collecting data from different ages in life-span studies means that ages are usually grouped into 5- or 10-year bins. The availability of internet-based psychological tests has increased sample size potential, allowing for finer-grain analysis of age effects (e.g., Hartshorne & Germine, 2015; Reimers & Maylor, 2005). The availability of large samples offsets the variability inherent in unsupervised testing (Chetverikov & Upravitelev, 2016; Enochson, & Culbertson, 2015; Hilbig, 2016).

Large samples from online data collection have also enabled more advanced analyses sensitive to age-related change. A seminal study examined the life-span trajectory of sustained attention using segmented regression, an analysis that estimates the rate of change and transition periods where the rate shifts (breakpoints; Fortenbaugh et al., 2015). The approach enabled them to find differential patterns for components of sustained attention: a criterion measure (responding during uncertainty) became increasingly conservative in a linear fashion with age, while a discrimination measure improved until mid-adulthood and then declined.

The goal of the current study was to perform a fine-grained examination of age-related differences using a web-based cognitive assessment of spatial working memory, interference control, processing speed, associative recognition, and set shifting (Cogniciti's Brain Health Assessment; Troyer et al., 2014). The Brain Health Assessment is well suited to examine cognitive aging as it was specifically designed for older adults, has high psychometric reliability and validity, and includes tasks sensitive to changes in the brain associated with aging and age-related cognitive disorders.

In addition to examining trajectories of age-related cognitive decline using measures of central tendency, we examined variability of performance between people (interindividual variability or diversity) and within a person across tasks (intraindividual variability or dispersion).

Diversity and dispersion both increase with age, even after accounting for age-related changes on mean performance (Christensen et al., 1994, 1999; Dykiert et al., 2012; Hilborn et al., 2009; Morse, 1993; Schretlen et al., 2003). Finally, we examined differences between age-related effects on each task, and whether all tasks loaded onto a single general factor, because general influences account for substantial variance in age-related cognitive decline (Salthouse, 2017; Salthouse & Meinz, 1995; Tucker-Drob et al., 2019; Verhaeghen, 2011).

Method

Participants

Participants were recruited via in-person brain health workshops, advertisements, media outlets, and word of mouth. The assessment was completed 115,973 times between 2014 and 2019 in individuals aged 14 to older than 100. Subsequent completions by the same individual were removed ($n = 20,687$). Completions were excluded if participants refreshed the page during the task ($n = 113$), had technical issues with data recording ($n = 15,541$), reported demographic information that was unclassifiable ($n = 26$), or reported health conditions that could affect cognitive performance (e.g., alcohol or substance abuse, stroke, traumatic brain injury, cancer treated with chemotherapy; $n = 7,168$).

The age range was narrowed to 18–90 ($n = 564$ removed), due to smaller samples (less than 50 per age), and extremely variable responses for ages outside this range. Completions were excluded if participants reported an age that did not appear to be accurate either due to disproportionate numbers of reported ages at the cutoffs of the validated range ($n = 8,093$), or if there was a jump in the reported age from one completion to another ($n = 4,078$). An assessment score was provided only to ages 50–79 in the first few years, ages 40–79 in the mid-years, and eventually for almost all ages, that is, 20–94; nearly twice as many individuals reported their age at the cutoffs of the validated age range (“40,” “50,” or “79”) compared to nearby ages, and their performance deviated from others of similar ages, suggesting that some had falsely reported their age to get a score. Data were also excluded if a participant's age changed by over 5 years, because data collection occurred only over 5 years (5 years was selected to allow for genuine mistakes, while excluding deliberate misinformation).

The final sample after exclusions consisted of 59,703 individuals. The mean age was 62.9 years ($SD = 11.8$; median = 64, interquartile range = 14), and 66% identified as female. Most participants had completed university (6% did not complete high school, 28% completed high school, 44% completed a college/undergraduate degree, 22% had a graduate/professional degree). Refer to [Supplementary Table S1](#) for sample size and demographic information per age decade.

Brain Health Assessment

The Brain Health Assessment is a free self-administered online test designed for older adults concerned about their memory. Full details of the tasks and development can be accessed in the original paper (Troyer et al., 2014). Participants completed the test in their own homes, by accessing the Cogniciti website (<https://cogniciti.com/>). The test takes around 20 min to complete. It consists of a background questionnaire and four cognitive tasks. The background questionnaire includes details of participants' age, sex, level of education, and specific health conditions that may affect cognition (e.g., high cholesterol, Alzheimer's disease, anxiety, insomnia or other sleep disorders, diabetes, stroke).

The four cognitive tasks were administered in the following order. A Spatial Working Memory task required participants to find pairs of shapes on a grid by remembering their locations over two trials. A number–word Stroop task, a measure of interference control, required participants to indicate the number of words for neutral (e.g., “boy boy”), congruent (e.g., “two two”), or incongruent stimuli (e.g., “one one”). A Face–Name Association task, a measure of associative memory, required participants to learn pairs of faces and names. A Letter–Number Alternation task, a measure of set shifting and attention control, required participants to click alternating numbers in ascending order. Refer to [Supplementary Material](#) for additional task details.

Tests were designed and selected for sensitivity to aging and ability to be completed online. Extensive development and piloting were conducted to ensure that the tasks could be completed reliably by people with basic computer skills. Practice trials with feedback are provided for the Stroop and Letter–Number Alternation tasks. Visual examples are provided for all tasks. After completing the test, participants received an overall score, which was a percentile value of their performance based on norms for age.

Measures

Total score

A single representative measure per task was used to calculate the total score: the total number of clicks for the Spatial Working Memory task (combining Trials 1 and 2), the response time on the incongruent condition for the Stroop task (because it combines baseline speed and interference effects), the associative recognition rate for the Face–Name Association task, and the completion time for the Letter–Number Alternation task. Scores per task were converted to z scores so that all tasks were in the same metric, and averaged to create the total score. Scores from the Face–Name Association task were

reversed so that in all tasks, higher scores reflected worse performance.

Individual task scores

Performance on the Spatial Working Memory task was measured as the total number of clicks on Trials 1 and 2. Stroop performance was measured as the average response time of correct responses in the Neutral condition (to measure baseline speed), and as the residuals after regressing Neutral response times from the Incongruent condition (to measure interference effects) and the Congruent condition (to measure facilitation effects; MacLeod, 1991). Residual scores control for variance due to baseline speed (Salthouse & Meinz, 1995). On speeded tasks (Letter–Number Alternation and Stroop), accuracy was also included as a measure of performance.

Process dissociation logic (Jacoby, 1991) was applied to the Face–Name Association task, to parse item recognition memory (I) from associative recognition memory (A; see Troyer et al., 2012). Hits to intact items involve identifying associations between items (A) and individual items in the absence of associations ($I[1 - A]$). Thus, hits to intact face–name pairs reflect both associative and item memory ($\text{yesIntact} = A + I[1 - A]$). False alarms to recombined face–name pairs reflect item memory only ($\text{yesRecombined} = I[1 - A]$), in the absence of associations. Thus, associative memory can be calculated as the difference between the proportions of hits to intact pairs and false alarms to recombined pairs ($\text{yesIntact} - \text{yesRecombined}$).

Variability

Diversity and dispersion were measured after controlling for confounding effects of background variables and age. These purification steps resulted in residual scores uncontaminated by age differences or background factors that influence performance (Hultsch et al., 2002). The residual scores were then used to calculate variability.

Accounting for mean performance is important because more extreme means are also associated with more variance (e.g., on the Stroop task, mean performance increases with age, and larger means are associated with greater variability). Substantial age-related increases in variability have been found, even after controlling for mean effects of age (e.g., Hultsch et al., 2002, 2008). To fully account for mean age-related effects, the purified residual scores were obtained from the final selected best-fitting model (refer to the subsection on *Age Differences in Mean Performance*), thus removing all effects of age on mean performance.

For diversity measurements, the purified residuals per task were converted to absolute scores to remove negative values. The absolute residual scores for each task were then regressed on age using the final selected best-fitting model. Low values reflect similar performance between different individuals, while high values reflect a large range in performance.

For dispersion measurements, the purified residual scores per task were converted to z scores, to allow comparisons across tasks. To calculate a single dispersion score, we used the same representative measure per task as the total score. Intraindividual standard deviations were calculated for each individual across the residual z scores for each task. The intraindividual standard deviation of standardized purified residuals was then regressed on age using the final selected best-fitting model. Low values reflect relatively similar performance (i.e., low variability) across tasks for an individual, while high values reflect uneven performance across tasks.

Data Preparation

Responses were excluded for practice trials and the first test trial on all tasks. Data were removed on tasks that had incomplete responses, but the participants were retained for other tasks if their data were complete ($n = 9,038$ to $n = 12,318$). Data were then trimmed per task, in an iterative manner, with a recursive moving criterion for the standard deviation based on the sample size (Grange, 2015). Within-subject trimming (per participant per condition) was done on individual trials for the Stroop task, to remove exceptionally low or high response times. A lower bound for legitimate responses was set at 250 ms. Within-subject trimming was not performed on other tasks as they did not contain extreme responses per trial (the range was from 0% to 100%; Face–Name Association task), or the measure of interest was the total score and not average performance across trials (the Letter–Number Alternation task and the Spatial Working Memory task). Between-subject trimming (per age per condition) was done on all tasks to remove individuals with exceptionally low or high scores ($n = 369$ to $n = 1,871$ removed per task; 1%–4% of data). The final samples were $n = 48,510$ for the Spatial Working Memory task, $n = 45,774$ for the Face–Name Association task, $n = 47,375$ for the Stroop task, and $n = 43,298$ for the Letter–Number Alternation task.

Analyses

Analyses were conducted with the R language and environment for statistical computing (R Core Team, 2020), using packages for data trimming (Grange, 2015), segmented regression (Muggeo, 2008), and confirmatory factor analysis (Rosseel, 2012).

Age differences

Locally estimated scatterplot smoothing.—Nonparametric locally estimated scatterplot smoothing (LOESS; Cleveland & Devlin, 1988) was used to visually examine performance at different ages. LOESS curves fit models to localized subsets in the data, using a span of $\alpha = .5$ to foresee trends over the means for each age.

Segmented regression.—Parametric regression models were used to quantify the pattern of change with age observed with the LOESS curves. We used segmented regression as it quantifies parameters of interest: the rate of change per year of age (the slope parameter, β), and the age(s) in which the rate shifts (the breakpoint parameter, which estimates statistically significant slope changes in the regression line, ψ).

Linear and polynomial regression.—We also fitted models with linear and polynomial effects of age, because these have been widely used in past studies to model age-related change.

Sample size.—Simulation studies have proposed a metric of $n = 500$ per breakpoint across a range of error precision scenarios, and a metric of $n = 1,000$ per sample across a range of breakpoint locations and slope coefficients (White et al., 2018). This offers a probability that the true parameter is included in over 90% of samples (Muggeo, 2003). The sample contained at least $n = 1,000$ per 10 years of age (ages 20–29, 30–39, etc.) allowing estimation of one breakpoint per decade.

The sample size was not evenly distributed across ages, with greater sample sizes for the age range of 50–80 than other ages (refer to Figure 1 and Supplementary Table S1), probably because the test was targeted at adults concerned about their memory as they aged, and because in the earlier years of measurement, validated scores were only provided to adults between 50 and 79. However, ages with fewer individuals still contained over 1,000 participants per decade.

Segmented model fitting.—To ensure robust estimates, 50 bootstrap samples were calculated, with a convergence tolerance of 0.00001, and a maximum of 20 iterations. Models were fit with no starting values for the breakpoints. To examine the reliability of estimates, models were refitted using starting values statistically calculated by a Davies test for a nonzero difference in the slope parameter (Davies, 2002). In all cases, no significant difference was found between the models fitted with and without prespecified initial estimates, $ps > .05$.

Increasingly complex models were fitted for each measure, beginning with the linear model (a no-breakpoint model), followed by a model in which the line was allowed to shift at one point (a one-breakpoint model), then a model in which the line was allowed to shift at two points (a two-breakpoint model), and so on. Increasingly complex models were compared using significance testing (hierarchical regression with a chi-squared difference test) and model parsimony criteria (the Akaike information criterion [AIC] and the Bayesian information criterion [BIC]). When criteria diverged, the best-fitting model was selected using the most conservative criterion, given that overfitting is a concern with very large data sets.

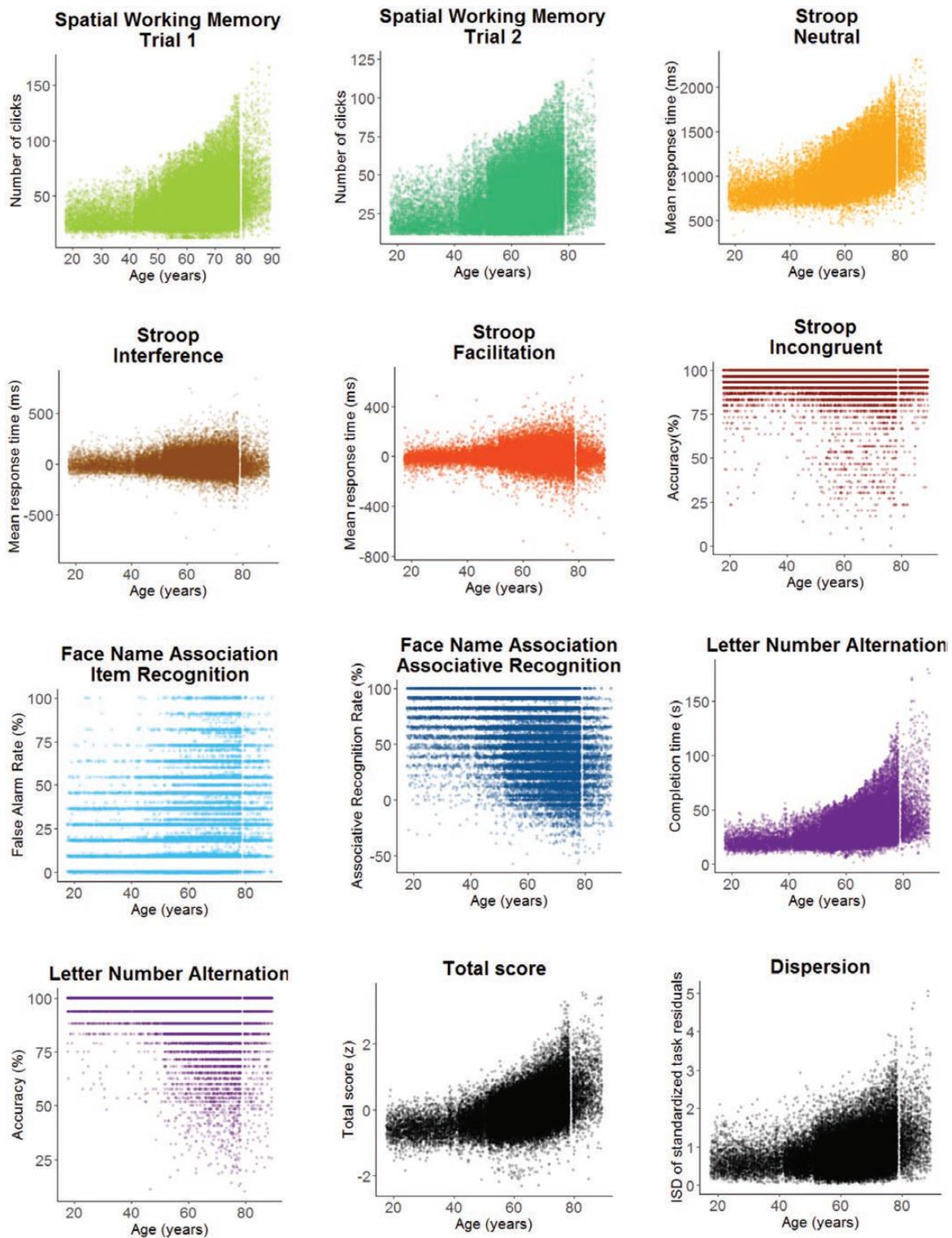


Figure 1. Density plots of performance per task.

Manipulation check for sampling bias (equal samples data set).— A manipulation check was run with equal samples per year of age, to account for the uneven distribution of participants across ages, and test whether breakpoints were influenced by the increased power for some ages. Random samples were taken from the complete data to

create a subset with $n = 50$ per age (as there were at least 50 individuals for each year of age).

Models were fit to the equal samples subset in the same way as the complete data set. When the best-fitting model diverged between the equal samples and complete data set, results from the equal samples data set were used to select

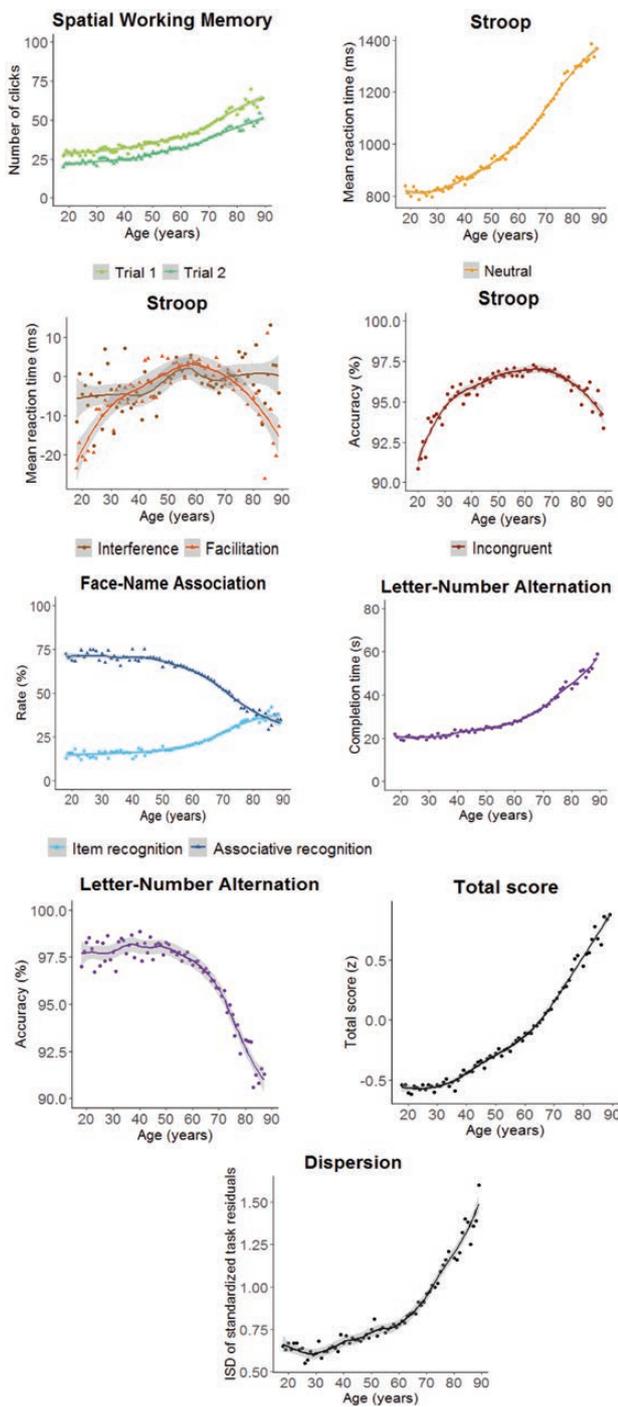


Figure 2. Locally estimated scatterplot smoothing (LOESS) curves of performance per task. *Note:* Each dot shows the mean performance per age. The gray shading around the LOESS curve indicates a 95% CI envelope.

the best-fitting model, because this data set removes bias from uneven sampling. The selected model from the equal samples data set was then fitted to the complete data set to obtain model estimates. This technique removed bias from uneven sampling across ages, yet used the maximal power from the whole sample.

Model comparison and selection.—The linear, polynomial, and best-fitting segmented regression models were compared, using goodness-of-fit (the percentage of variance explained, R^2 ; and the residual standard error) and parsimony criteria (AIC, BIC). The models were not nested and thus cannot be statistically compared to each other.

Estimates of the parameters of interest are presented for the final selected model per task. The regression coefficients (slope estimates) estimate the amount of change per year of age, while the breakpoint locations estimate the ages during which a shift in performance occurs. Parameter estimates are presented using raw effect sizes, because these are intuitive to interpret as they use the original units of performance.

Comparison of performance between tasks

Representative z scores per task (the same scores as used in the total score) were used to compare between tasks. Scores from the Face–Name Association task were reversed so that for all tasks, higher scores reflected worse performance.

Standardized metric for each task.—The final selected models for each task were compared to examine performance across tasks. Comparisons between tasks were done with visual examination and effect sizes rather than significance tests, as large samples produce highly precise estimates and low p values for all effects tested, regardless of size, theoretical significance, or importance (Fan et al., 2021).

General factor across tasks.—Confirmatory factor analysis was used to test how well performance on individual task scores loaded onto a single factor. Robust maximum likelihood estimation was used to fit the model to sample covariances, which is ideal because data for each task were continuous and on the same scale. Model fit was assessed with the chi-squared test statistic (χ^2), comparative fit index (CFI: $\geq .90$ is good and $\geq .95$ is excellent), Tucker–Lewis index (TLI: $\geq .90$ is good and $\geq .95$ is excellent), root mean square error of approximation (RMSEA: < 0.08 is reasonable and < 0.05 is good), and standardized root mean square residual (SRMR: < 0.08 is good; Kenny, 2015). In addition, effects of age on each task were recalculated after accounting for the general factor of variance shared across tasks, to parse general and specific influences (refer to [Supplementary Material](#)).

Results

Density plots of the raw data are presented in [Figure 1](#).

Age Differences in Mean Performance

LOESS curves of the total score and individual tasks indicated that performance very slightly declined over early and middle adulthood, followed by a more rapid decline

Table 1. Comparison of Linear, Polynomial, and Segmented Regression Models

Task	Measure (units)	Linear model	Polynomial	Segmented
Spatial Working Memory	Trial 1 (number of clicks)	$R^2 = 0.14$, RSE = 17.0 AIC = 400967, BIC = 401222	$R^2 = 0.16$, RSE = 16.9 AIC = 400185, BIC = 400448	$R^2 = 0.16$, RSE = 16.9 AIC = 400114, BIC = 400385
	Trial 2 (number of clicks)	$R^2 = 0.14$, RSE = 14.5 AIC = 386110, BIC = 386364	$R^2 = 0.15$, RSE = 14.5 AIC = 385719, BIC = 385982	$R^2 = 0.15$, RSE = 14.5 AIC = 385685, BIC = 385957
Stroop	Neutral RT (ms)	$R^2 = 0.32$, RSE = 179 AIC = 612702, BIC = 612990	$R^2 = 0.34$, RSE = 177 AIC = 611522, BIC = 611819	$R^2 = 0.34$, RSE = 177 AIC = 611519, BIC = 611825
	Interference RT (ms)	$R^2 = 0.003$, RSE = 82.5 AIC = 535865, BIC = 536013	$R^2 = 0.003$, RSE = 82.5 AIC = 535866, BIC = 536024	$R^2 = 0.003$, RSE = 82.5 AIC = 535866, BIC = 536032
	Facilitation RT (ms)	$R^2 = 0.003$, RSE = 61.4 AIC = 508677, BIC = 508956	$R^2 = 0.006$, RSE = 61.3 AIC = 508542, BIC = 508830	$R^2 = 0.006$, RSE = 61.3 AIC = 508547, BIC = 508844
	Accuracy (%)	$R^2 = 0.007$, RSE = 6.69 AIC = 310340, BIC = 310602	$R^2 = 0.02$, RSE = 6.66 AIC = 309813, BIC = 310084	$R^2 = 0.02$, RSE = 6.66 AIC = 309905, BIC = 310185
Face-Name Association	Item recognition (%)	$R^2 = 0.09$, RSE = 19.3 AIC = 404014, BIC = 404268	$R^2 = 0.10$, RSE = 19.2 AIC = 403612, BIC = 403874	$R^2 = 0.10$, RSE = 19.2 AIC = 403567, BIC = 403838
	Associative recognition (%)	$R^2 = 0.14$, RSE = 26.1 AIC = 431920, BIC = 432200	$R^2 = 0.15$, RSE = 26.0 AIC = 431345, BIC = 431634	$R^2 = 0.15$, RSE = 25.9 AIC = 431298, BIC = 431596
Letter-Number Alternation	Completion time (seconds)	$R^2 = 0.22$, RSE = 12.5 AIC = 337006, BIC = 337283	$R^2 = 0.25$, RSE = 12.3 AIC = 335450, BIC = 335736	$R^2 = 0.25$, RSE = 12.3 AIC = 335393, BIC = 335688
	Accuracy (%)	$R^2 = 0.035$, RSE = 8.83 AIC = 307258, BIC = 307535	$R^2 = 0.043$, RSE = 8.79 AIC = 306921, BIC = 307207	$R^2 = 0.043$, RSE = 8.79 AIC = 306911, BIC = 307206
Total	Mean across tasks (z)	$R^2 = 0.26$, RSE = 0.47 AIC = 53839, BIC = 54123	$R^2 = 0.28$, RSE = 0.46 AIC = 52808, BIC = 53102	$R^2 = 0.28$, RSE = 0.46 AIC = 52814, BIC = 53116
All tasks	Dispersion	$R^2 = 0.12$, RSE = 0.42 AIC = 44993, BIC = 45277	$R^2 = 0.13$, RSE = 0.42 AIC = 44330, BIC = 44623	$R^2 = 0.13$, RSE = 0.42 AIC = 44255, BIC = 44557
All tasks	General factor	$R^2 = 0.37$, RSE = 0.67 AIC = 83531, BIC = 83746	$R^2 = 0.40$, RSE = 0.65 AIC = 81341, BIC = 81350	$R^2 = 0.41$, RSE = 0.65 AIC = 81351, BIC = 815583

Note: AIC = Akaike information criterion; BIC = Bayesian information criterion; R^2 = percentage of variance explained; RSE = residual standard error; RT = response time. The best-fitting model is highlighted in bold.

beginning around age 60 (Figure 2). The pattern of decline later in life was observed across all measures with the exception of interference response times on the Stroop task. There appeared to be a tradeoff between speed and accuracy on the Stroop task, because speed declined over adulthood and more rapidly after the 60s, while accuracy improved over adulthood and then declined slightly after the 60s.

The best-fitting segmented regression model had one breakpoint for all measures, with the exception of the interference effect on the Stroop task which was best fit by a linear model. Results from our manipulation check on the equal samples data set revealed that the parameter estimates from the equal samples data set were similar to those from the complete data set (refer to Supplementary Tables S2–S4 for complete results on model comparison for the equal samples and complete data sets).

Segmented and polynomial models both fit the data better than linear models, and had nearly identical fits to each other (Table 1). Both had the same values for the amount of variance explained (R^2) and the average residual standard error. The segmented model performed slightly better on parsimony measures (AIC and BIC): the

segmented model offered a more parsimonious fit on seven measures and the polynomial model on three measures, and the models were equal on one measure.

In line with past recommendations where segmented and polynomial models provided similar fits (Ryan & Porth, 2007), we selected the segmented model as our final model because it quantifies the age at which performance shifts. In contrast, the maximum/minimum points from polynomial regression may not be relevant to interpret, because no maximum or minimum value is expected for the age range.

The transition points and rates of change per age are described for the final selected model per measure in Table 2. For example, on the total score, there was a gradual decline in performance of 0.01 units per year from the start of measurement at age 18 until 62.5 years, followed by a more rapid decrease of 0.04 units per year until the end of measurement at age 90: 95% CIs: [0.01, 0.02], [61.9, 63.2], and [0.03, 0.04], respectively. Overall, performance on individual tasks and the total score showed a gradual decline from the start of measurement at age 18 until around age 60, followed by a more rapid decline until the end of measurement at age 90 (Figure 3).

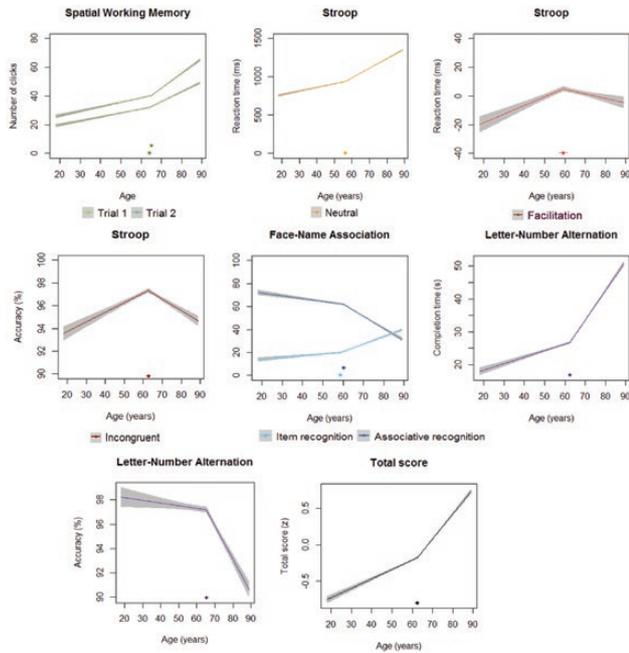


Figure 3. Segmented regression models of mean performance per task. *Note:* Breakpoints [95% CI] are shown along the y-axis. Models shown are fitted to the complete data set, selected using the best-fitting model from the equal samples data set.

Age Differences in Diversity Across People

Density plots of the raw scores reflect an increase in variability between people as a function of age (Figure 1). A Levene’s test for homogeneity of variance was conducted on standard deviations for different ages to examine if there was diversity in the data. The test was significant for all ages on each task, $p < .05$, indicating an age-related increase in variability among individuals.

The best-fitting segmented regression model had one breakpoint for each measure (Table 2; Figure 4). Diversity gradually increased from the start of measurement at age 18 until around age 60, followed by a more rapid increase until the end of measurement at age 90. The exception was the Face-Name Association task, which showed no effects of age on diversity for both measures (item and associative recognition).

Age Differences in Dispersion Across Tasks

A LOESS curve indicated a slight decrease in dispersion from age 18 until age 30 followed by a gradual increase in dispersion until around age 60, and then a rapid increase in dispersion until age 90 (Figure 1). The best-fitting segmented regression model had one breakpoint (Figure 4). Dispersion gradually increased by 0.01 units from age 18 until age 64.7, followed by a more rapid increase of 0.02 units until age 90: 95% CIs: [0.005, 0.01], [63.9, 65.4], and [0.02, 0.03], respectively.

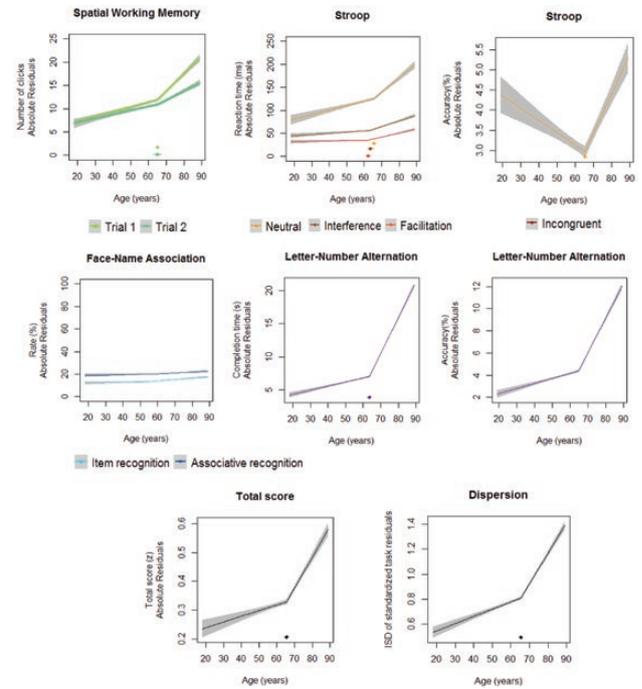


Figure 4. Segmented regression models of diversity per task, and dispersion across tasks. *Note:* Breakpoints [95% CI] are shown along the y-axis. Models shown are fitted to the complete data set, selected using the best-fitting model from the equal samples data set.

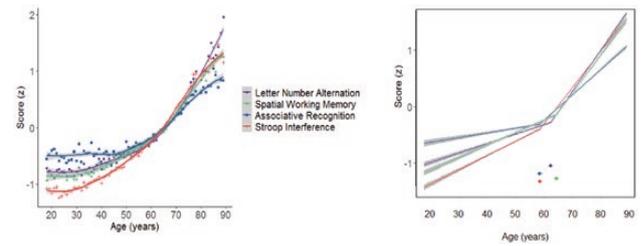


Figure 5. Locally estimated scatterplot smoothing curves and segmented regression models of performance between tasks (using a single representative measure per task, on a standardized scale; z scores). *Note:* The direction of scores was reversed for the Face-Name Association task so that higher scores represent worse performance on all tasks.

Comparison of Performance Between Tasks

Standardized metric for each task

LOESS curves indicated that the pattern of change across tasks was very similar when standardized scores were compared (Figure 5). Segmented regression also showed that the rates and breakpoints for each task were very similar. Performance between tasks became even more similar between people after the breakpoint in the 60s.

General factor across tasks

Confirmatory factor analysis revealed that a single-factor model offered a good fit to the data, with an excellent CFI

Table 2. Transition Ages and Slopes (Rate of Change per Year [95% CI]) on the Final Selected Model of Mean Performance and Diversity Across People for the Online Participants (Aged 18–90)

Task	Measure (units)	Mean performance			Diversity		
		Adulthood slope	Age of transition	Older adulthood slope	Adulthood slope	Age of transition	Older adulthood slope
Spatial Working Memory	Trial 1 (number of clicks)	0.3 [0.3, 0.3]	65.1 [64.4, 65.8]	1.0 [1.0, 1.1]	0.1 [0.09, 0.13]	65.1 [63.8, 66.3]	0.4 [0.3, 0.4]
	Trial 2 (number of clicks)	0.3 [0.3, 0.3]	64.1 [63.1, 65.1]	0.7 [0.70, 0.74]	0.08 [0.06, 0.1]	65.0 [62.8, 67.2]	0.2 [0.2, 0.2]
Stroop	Neutral RT (ms)	4.9 [4.2, 5.2]	55.7 [54.9, 56.4]	12.5 [12.0, 12.9]	1.0 [0.7, 1.2]	65.6 [64.2, 67.1]	3.2 [2.8, 3.6]
	Interference RT (ms)	0.02 [-0.1, 0.1]			0.3 [0.1, 0.4]	63.5 [61.9, 65.1]	1.3 [1.1, 1.5]
	Facilitation RT (ms)	0.6 [0.5, 0.7]	59.2 [57.2, 61.3]	-0.3 [-0.5, -0.2]	0.07 [-0.03, 0.2]	62.4 [60.8, 63.9]	0.9 [0.8, 1.0]
	Incongruent accuracy (%)	0.08 [0.07, 0.1]	62.8 [61.8, 63.8]	-0.09 [-0.1, 0.07]	-0.03 [-0.04, -0.02]	65.4 [64.2, 66.5]	0.1 [0.08, 0.1]
Face-Name Association	Item recognition (%)	0.2 [0.1, 0.2]	58.7 [57.5, 59.9]	0.7 [0.6, 0.7]	0.08 [0.06, 0.1]		
	Associative recognition (%)	-0.2 [-0.3, -0.2]	60.4 [59.4, 61.3]	-1.1 [-1.1, -1.0]	0.05 [0.02, 0.08]		
Letter-Number Alternation	Completion time (seconds)	0.3 [0.2, 0.3]	62.5 [62.0, 63.1]	1.0 [0.9, 1.0]	0.1 [0.05, 0.07]	63.6 [63.1, 64.1]	0.56 [0.5, 0.6]
	Accuracy (%)	-0.04 [-0.1, -0.02]	65.3 [64.2, 66.4]	-0.3 [-0.33, -0.26]	0.05 [0.04, 0.05]	64.8 [64.0, 65.5]	0.3 [0.29, 0.33]
All tasks	Mean across tasks (z)	0.01 [0.01, 0.02]	62.6 [61.9, 63.2]	0.04 [0.03, 0.04]	0.002 [0.001, 0.003]	65.5 [64.5, 66.6]	0.01 [0.009, 0.01]

Note: The slopes can be interpreted as the number of units a measure changes per year of age.

of .98, an excellent TLI of .94, an acceptable RMSEA of .08, and a good SRMR of .03. All tasks showed good loadings onto the general factor (>.05). This suggests that considerable variance is shared across the tasks. Further analyses parsing general and specific influences showed that for each task, unique effects of age were found even after accounting for the general factor of variance shared across tasks from (refer Supplemental Table S5).

Discussion

The prototypical cognitive aging profile is held to be linear or polynomial change over the life span, for mean performance, diversity, and dispersion (Borella et al., 2008; Christensen et al., 1994, 1999; Hultsch et al., 2002; Myerson et al., 2003; Salthouse & Meinz, 1995; Salthouse, 2017; Uttl & Graf, 1997). Leveraging a large sample size collected online over 5 years, we modeled how and when fluid cognitive abilities vary with age at a finer grain than previous studies. We found a non-linear aging pattern, with gradual decline from ages 18 until the early 60s, followed by rapid decline until age 90. The observed pattern was found on spatial working memory, processing speed, associative recognition, and

set shifting, but not on interference control. Diversity across people and dispersion across tasks also gradually increased from 18 until the early 60s, and then rapidly increased until age 90. Our findings indicate that mean performance declines with advancing age, but that relatively more age-related decline occurs for some people than others, and that relatively more age-related decline occurs on some tasks than others.

A strength of the current study was the use of a wide age range of individuals across the span of adulthood. Most studies of aging compare older and younger adult groups, leaving out midlife. Including middle-aged individuals offered us insights into the midlife period between younger and older adulthood, in which we observed small but present decline. Our finding of two stages of decline (one during adulthood and one later in life) highlights the value of studying age as a continuous variable and of considering stages in life-span aging (Craik & Bialystok, 2007).

Our results on increased variability with age highlight the importance of studying variability in addition to mean performance, to capture both the average pattern and deviations from this pattern. We found that a shift to increased variability occurs around the same ages as a shift to increased decline in mean performance, which is notable because variability was calculated after purification steps to

account for mean performance. The rise in diversity across people later in life is an issue for the study of mean performance, both statistically (because it influences fit statistics and underlying assumptions) and theoretically, but is interesting in itself. Future work is needed to explore the factors that produce the increase in diversity, such as socioeconomic status or health factors (e.g., vascular conditions).

The early 60s as a window signaling accelerated decline is applicable to the classification of older adulthood, which varies substantially across studies. Some researchers classify older adult groups beginning at 50, others at age 70. Moreover, they fail to justify their age selection. Our finding suggests that classifying individuals above 50 as older adults underestimates age-related decline on the abilities measured, while beginning classification over 70 misses early decline.

Past cognitive life-span data appear to also identify a transition around age 60, although a specific transition was not formally assessed (e.g., Myerson et al., 2003, Figure 3; Tombaugh, 2004, Figure 1; Uttl & Graf, 1997, Figure 1). The timing of the transition corresponds with a shift in late midlife to accelerated decline in aging of the prefrontal cortex (Cabr e et al., 2017; which mediates the attention and executive functioning tasks in our battery; Yuan & Raz, 2014) and in less segregation of brain systems (Chan et al., 2014; which is predictive of episodic memory, also measured in our battery). A possible reason is that considerably accelerated decline of some cognitive abilities occurs in the early to mid-60s after retirement Xue et al., 2018; although other factors also come into play, Denier et al., 2017. For example, episodic memory showed a twofold decline following retirement, even after adjusting for health, age, and wealth ($n = 18,575$; Clouston & Denier, 2017). However, future work is needed to establish causality.

The observed slope estimates are useful to estimate age differences by quantifying the magnitude of decline (i.e., the change per year of age). For example, on the Letter-Number Alternation task, adults completed the task 0.3 s slower per year of age from age 18 until the 60s, and then completed the task 1 s slower per year of age until age 90. Estimating each age as a timepoint is more informative than stating that older adults took x seconds longer to complete the task than young adults. Gradual changes, although small, can result in subtle changes in everyday skills (e.g., driving; Harada et al., 2013). Characterizing normal aging also provides a baseline for identifying abnormal trajectories.

Modeling Life-Span Age Differences

Given the large sample size, presenting the mean for each age and visualizing the pattern via LOESS curves is adequate to understand the pattern of age effects in a theoretically neutral data-driven manner. The LOESS curves are the most reliable of the different modeling approaches as they do not make any assumptions about the underlying distribution. Nevertheless, parametric

approaches are useful to quantify the nonlinear pattern of age effects, with the caveat of making distributional assumptions.

Our findings offer insights into the comparative fit between polynomial and segmented models. Both models had nearly identical goodness of fit. Polynomial models have precedence in the life-span literature, but our findings show that what had traditionally been perceived as gradual acceleration of cognitive decline is equally well modeled by two phases of linear decline.

The comparable fit makes final model selection ambiguous. On the one hand, polynomial models are more parsimonious as they have one parameter less. On the other hand, the additional parameter from segmented regression offers theoretically interesting information, delineating the age range when performance changes from gradual to accelerated decline. Further, for our sample, the segmented approach performed better on parsimony criteria for more measures than the polynomial approach, despite the additional parameter.

Our findings do not clarify whether the true underlying nonlinear pattern is curvilinear or transitional/piecewise. Caution is warranted against overinterpreting the transition timing without future work to clarify whether the observed breakpoints are replicated, and whether the statistical meaning that they represent (a significant change in the slope) holds psychological meaning. Future work could also explore segmented curvilinear functions (which were not used here as they are currently less accessible to fit and to interpret, while segmented regression offers a good tradeoff between model complexity and parameter interpretability; Cudeck & du Toit, 2002; Muggeo, 2003; Zapata, 2019).

Comparison Between Cognitive Abilities

Slightly diverging trajectories were observed on tasks with multiple conditions. On the Spatial Working Memory task, decline after the 60s was smaller on the second trial than the first trial, indicating a slight improvement in performance between the trials (possibly from learning or task familiarization). On the Face-Name Association task, with advancing age, performance relied less on associative recognition memory and more on item recognition memory, in line with the associative deficit hypothesis and past research that older adults may have a specific deficit on creating bindings between items and may rely more on familiarity with individual items to compensate for decline in remembering associations between items (Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008). On the Stroop task, a large effect of age was found on processing speed (the neutral condition), a small effect on facilitation (the congruent condition after controlling for processing speed), and no effect on interference (the incongruent condition after controlling for processing speed). Our finding of no effects of age on interference control in the Stroop task is in line with past findings and a recent meta-analysis

that found no general deficit on inhibition after accounting for processing speed, when measured by the Stroop task and other tests of inhibition (Rey-Mermet & Gade, 2018; Uttl & Graf, 1997). The results fail to support the inhibition deficit hypothesis of a specific age-related deficit on inhibiting distracting information (Chiappe et al., 2000; Hasher & Zacks, 1988). However, effects of age on inhibition are moderated by difficulty, and the number naming version of the Stroop that we used is easier than the standard color naming version (Bugg et al., 2007).

Our findings across tasks show that there is little difference in rates of change and shifts in performance between tasks, indicating that a general factor offers a parsimonious explanation for some of the observed effects of age on specific tasks. Further, all tasks loaded well onto a single general factor, which supports past findings that shared variance across cognitive tasks substantially accounts for effects of age on specific tasks (Salthouse, 2017; Verhaeghen, 2011). This is consistent with the dedifferentiation hypothesis that individual cognitive abilities become more correlated in old age (Balinsky, 1941; Baltes & Lindenberger, 1997). Effects of age on individual tasks were smaller (but still present) after controlling for an estimate of general influences derived from the other measures, which maps onto past work showing age differences on individual measures after dissociating general and specific influences (Salthouse, 2017). Our results align with past evidence that age influences on specific cognitive abilities cannot be accurately assessed without first accounting for general age influences (Salthouse & Meinz, 1995). Our finding indicates that effects of age across cognitive abilities occur along a common statistical dimension but does not indicate a single underlying cognitive construct or a single genetic or neurobiological cause (Craik et al., 2018; Tucker-Drob et al., 2019).

Limitations and Conclusions

A limitation of the current study is the use of an opportunity sample of individuals who took the online assessment on the test website. Participants were self-selected and motivated to complete the assessment. The assessment was specifically marketed for older adults with concerns about cognition, thus participants older than 60 may have joined because of cognitive concerns, while younger participants may have joined out of interest in their own cognition. The older participants may be lower-functioning than a random sample (from having more memory concerns than average), or may be higher-functioning (from being more informed about memory, or having high computer literacy). Finally, the data were cross-sectional. Future work with longitudinal data is required to parse out cohort effects.

Overall, the current findings show a pattern of gradual to rapid decline over adulthood, with a transition around the 60s, on cognitive tasks that were designed for their

sensitivity to aging and neurodegeneration. Our results reflect the utility of online assessments for rapid and reliable cognitive screening, as well as for studying cognitive aging in large groups.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* online.

Funding

This work was supported by research grants from the Natural Sciences and Engineering Council (NSERC) of Canada (grant number RGPIN-2017-06057) to N. D. Anderson and an Alzheimer Society of Canada (fellowship number 20-16) to A. LaPlume.

Conflict of Interest

None declared.

Acknowledgments

We thank Dr. S. Gardner for statistical consulting, and individuals who contributed to test creation or database management (Dr. Lynn Hasher, Dr. Gillian Rowe, Dr. Kelly Murphy, Dr. Larry Leach, Dr. Malcolm Binns, Michael Meagher, and Caitlin Johnston). Aggregate data, analytical code, and study materials will be made available upon request. The study was approved by the Baycrest Research Ethics Board (REB# 15-05). The research aims and analysis plan were preregistered prior to statistical analyses (but after data were collected) at https://osf.io/d8t3y/?view_only=0bb90234e104480e8f9780215642bafa.

Author Contributions

A. LaPlume developed the study concept and drafted the manuscript. Testing materials were created by B. Levine and A. K. Troyer, along with a team of researchers (c.f. Troyer et al., 2014). Data collection was coordinated by B. Levine and A. K. Troyer. Database access was coordinated by L. McKetton. A. LaPlume performed the data cleaning and preprocessing, analysis, and interpretation, under the supervision of N. D. Anderson, A. K. Troyer, and B. Levine. N. D. Anderson, A. K. Troyer, B. Levine, and L. McKetton provided critical revisions. All authors contributed to the study design and approved the final version of the manuscript for submission.

References

- Balinsky, B. (1941). An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs*, 23, 191–234.

- Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12*, 12–21. doi:10.1037/0882-7974.12.1.12
- Bender, A. R., Naveh-Benjamin, M., & Raz, N. (2010). Associative deficit in recognition memory in a lifespan sample of healthy adults. *Psychology and Aging, 25*(4), 940–948. doi:10.1037/a0020595
- Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 60*(5), 223–233. doi:10.1093/geronb/60.5.p223
- Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychologica, 128*(1), 33–44. doi:10.1016/j.actpsy.2007.09.008
- Bugg, J. M., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2007). Age differences in Stroop interference: Contributions of general slowing and task-specific deficits. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 14*(2), 155–167. doi:10.1080/138255891007065
- Cabré, R., Naudí, A., Dominguez-Gonzalez, M., Ayala, V., Jové, M., Mota-Martorell, N., Piñol-Ripoll, G., Gil-Villar, M. P., Rué, M., Portero-Otín, M., Ferrer, I., & Pamplona, R. (2017). Sixty years old is the breakpoint of human frontal cortex aging. *Free Radical Biology and Medicine, 103*, 14–22. doi:10.1016/j.freeradbiomed.2016.12.010
- Chan, M. Y., Park, D. C., Savalia, N. K., Petersen, S. E., & Wig, G. S. (2014). Decreased segregation of brain systems across the healthy adult lifespan. *Proceedings of the National Academy of Sciences of the United States of America, 111*(46), E4997–E5006. doi:10.1073/pnas.1415122111
- Chetverikov, A., & Upravitelev, P. (2016). Online versus offline: The web as a medium for response time data collection. *Behavior Research Methods, 48*(3), 1086–1099. doi:10.3758/s13428-015-0632-x
- Chiappe, P., Hasher, L., & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition, 28*(1), 8–17. doi:10.3758/bf03211570
- Christensen, H., Mackinnon, A., Jorm, A. F., Henderson, A. S., Scott, L. R., & Korten, A. E. (1994). Age differences and interindividual variation in cognition in community-dwelling elderly. *Psychology and Aging, 9*(3), 381–390. doi:10.1037//0882-7974.9.3.381
- Christensen, H., Mackinnon, A. J., Korten, A. E., Jorm, A. F., Henderson, A. S., & Jacomb, P. (1999). Dispersion in cognitive ability as a function of age: A longitudinal study of an elderly community sample. *Aging, Neuropsychology, and Cognition, 6*(3), 214–228. doi:10.1076/anec.6.3.214.779
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*(403), 596–610. doi:10.1080/01621459.1988.10478639
- Clouston, S. A., & Denier, N. (2017). Mental retirement and health selection: Analyses from the US Health and Retirement Study. *Social Science & Medicine, 178*, 78–86. doi:10.1016/j.socscimed.2017.01.019
- Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences, 10*(3), 131–138. doi:10.1016/j.tics.2006.01.007
- Craik, F. I. M., & Bialystok, E. (2007). Lifespan cognitive development: The roles of representation and control. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 557–601). Psychology Press.
- Craik, F. I. M., Eftekhari, E., Bialystok, E., & Anderson, N. D. (2018). Individual differences in executive functions and retrieval efficacy in older adults. *Psychology and Aging, 33*(8), 1105–1114. doi:10.1037/pag0000315
- Cudeck, R., & du Toit, S. H. (2002). A version of quadratic regression with interpretable parameters. *Multivariate Behavioral Research, 37*(4), 501–519. doi:10.1207/S15327906MBR3704_04
- Davies, R. B. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika, 89*, 484–489. <http://www.jstor.org/stable/4140594>
- Denier, N., Clouston, S. A., Richards, M., & Hofer, S. M. (2017). Retirement and cognition: A life course view. *Advances in Life Course Research, 31*, 11–21. doi:10.1016/j.alcr.2016.10.004
- Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Age differences in intra-individual variability in simple and choice reaction time: Systematic review and meta-analysis. *PLoS One, 7*(10), e45759. doi:10.1371/journal.pone.0045759
- Enochson, K., & Culbertson, J. (2015). Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLoS One, 10*, e0116946. doi:10.1371/journal.pone.0116946
- Fan, C. L., Abdi, H., & Levine, B. (2021). On the relationship between trait autobiographical episodic memory and spatial navigation. *Memory & Cognition, 49*(2), 265–275. doi:10.3758/s13421-020-01093-7
- Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological Science, 26*(9), 1497–1510. doi:10.1177/0956797615594896
- Grange, J. A. (2015). *trimr: An implementation of common response time trimming methods. R package version 1.0.1.* <https://CRAN.R-project.org/package=trimr>
- Harada, C. N., Love, M. C. N., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine, 29*(4), 737–752. doi:10.1016/j.cger.2013.07.002
- Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26*(4), 433–443. doi:10.1177/0956797614567339
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 193–225). Academic Press.
- Hilbig, B. E. (2016). Reaction time effects in lab- versus web-based research: Experimental evidence. *Behavior Research Methods, 48*(4), 1718–1724. doi:10.3758/s13428-015-0678-9
- Hilborn, J. V., Strauss, E., Hultsch, D. F., & Hunter, M. A. (2009). Intraindividual variability across cognitive domains: Investigation of dispersion levels and performance profiles in older adults.

- Journal of Clinical and Experimental Neuropsychology*, 31(4), 412–424. doi:10.1080/13803390802232659
- Hultsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 57(2), 101–115. doi:10.1093/geronb/57.2.p101
- Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. (2008). Intraindividual variability, cognition, and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed., pp. 491–556). Psychology Press.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541. doi:10.1016/0749-596X(91)90025-F
- Kenny, D. A. (2015, November 24). *Measuring model fit*. <http://davidakenny.net/cm/fit.htm>
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. doi:10.1037/0033-2909.109.2.163
- Morse, C. K. (1993). Does variability increase with age? An archival study of cognitive measures. *Psychology and Aging*, 8(2), 156–164. doi:10.1037//0882-7974.8.2.156
- Muggeo, V. M. (2008). Segmented: An R package to fit regression models with broken-line relationships. *R News*, 8(1), 20–25. <https://cran.r-project.org/doc/Rnews>
- Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19), 3055–3071. doi:10.1002/sim.1545
- Myerson, J., Emery, L., White, D. A., & Hale, S. (2003). Effects of age, domain, and processing demands on memory span: Evidence for differential decline. *Aging, Neuropsychology, and Cognition*, 10(1), 20–27. doi:10.1076/anec.10.1.20.13454
- Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1170–1187. doi:10.1037//0278-7393.26.5.1170
- Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, 23(1), 104–118. doi:10.1037/0882-7974.23.1.104
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, 17(2), 299–320. doi:10.1037/0882-7974.17.2.299
- Periáñez, J. A., Ríos-Lago, M., Rodríguez-Sánchez, J. M., Adrover-Roig, D., Sánchez-Cubillo, I., Crespo-Facorro, B., Quemada, J. I., & Barceló, F. (2007). Trail Making Test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Archives of Clinical Neuropsychology*, 22(4), 433–447. doi:10.1016/j.acn.2007.01.022
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Reimers, S., & Maylor, E. A. (2005). Task switching across the life span: Effects of age on general and specific switch costs. *Developmental Psychology*, 41(4), 661–671. doi:10.1037/0012-1649.41.4.661
- Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, 25(5), 1695–1716. doi:10.3758/s13423-017-1384-7
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5-12 (BETA). *Journal of Statistical Software*, 48(2), 1–36.
- Ryan, S. E. & Porth, L. S. (2007). *A tutorial on the piecewise regression approach applied to bedload transport data*. U.S. Department of Agriculture & Forest Service Rocky Mountain Research Station.
- Salthouse, T. A. (2017). Shared and unique influences on age-related cognitive change. *Neuropsychology*, 31(1), 11–19. doi:10.1037/neu0000330
- Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 50(6), 297–306. doi:10.1093/geronb/50b.6.p297
- Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M., & Hocking, K. J. (2000). Effects of aging on efficiency of task switching in a variant of the Trail Making Test. *Neuropsychology*, 14(1), 102–111. doi:10.1037/0894-4105.14.1.102
- Schretlen, D. J., Munro, C. A., Anthony, J. C., & Pearlson, G. D. (2003). Examining the range of normal intraindividual variability in neuropsychological test performance. *Journal of the International Neuropsychological Society*, 9(6), 864–870. doi:10.1017/S1355617703960061
- Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2), 203–214. doi:10.1016/S0887-6177(03)00039-8
- Troyer, A. K., Leach, L., & Strauss, E. (2006). Aging and response inhibition: Normative data for the Victoria Stroop Test. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 13(1), 20–35. doi:10.1080/138255890968187
- Troyer, A. K., Murphy, K. J., Anderson, N. D., Craik, F. I. M., Moscovitch, M., Maione, A., & Gao, F. (2012). Associative recognition in mild cognitive impairment: Relationship to hippocampal volume and apolipoprotein E. *Neuropsychologia*, 50(14), 3721–3728. doi:10.1016/j.neuropsychologia.2012.10.018
- Troyer, A. K., Rowe, G., Murphy, K. J., Levine, B., Leach, L., & Hasher, L. (2014). Development and evaluation of a self-administered on-line test of memory and attention for middle-aged and older adults. *Frontiers in Aging Neuroscience*, 6, 335. doi:10.3389/fnagi.2014.00335
- Tucker-Drob, E. M., Brandmaier, A. M., & Lindenberger, U. (2019). Coupled cognitive changes in adulthood: A meta-analysis. *Psychological Bulletin*, 145(3), 273–301. doi:10.1037/bul0000179
- Uttl, B., & Graf, P. (1997). Color-Word Stroop test performance across the adult lifespan. *Journal of Clinical and Experimental Neuropsychology*, 19(3), 405–420. doi:10.1080/01688639708403869
- Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science*, 20(3), 174–180. doi:10.1177/0963721411408772
- White, S. R., Muniz-Terrera, G., & Matthews, F. E. (2018). Sample size and classification error for Bayesian change-point

- models with unlabelled sub-groups and incomplete follow-up. *Statistical Methods in Medical Research*, 27(5), 1476–1497. doi:10.1177/0962280216662298
- Xue, B., Cadar, D., Fleischmann, M., Stansfeld, S., Carr, E., Kivimäki, M., McMunn, A., & Head, J. (2018). Effect of retirement on cognitive function: The Whitehall II cohort study. *European Journal of Epidemiology*, 33(10), 989–1001. doi:10.1007/s10654-017-0347-7
- Yuan, P., & Raz, N. (2014). Prefrontal cortex and executive functions in healthy adults: A meta-analysis of structural neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 42, 180–192. doi:10.1016/j.neubiorev.2014.02.005
- Zapata, A. (2019). *The ability of polynomial and piecewise regression models to fit polynomial, piecewise, and hybrid functional forms* [Doctoral dissertation, City University of New York]. CUNY Academic Works. https://academicworks.cuny.edu/gc_etds/3048