OXFORD

ORIGINAL ARTICLE

A Hierarchical Watershed Model of Fluid Intelligence in Childhood and Adolescence

Delia Fuhrmann, Ivan L. Simpson-Kent, Joe Bathelt, The CALM Team and Rogier A. Kievit

MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, CB2 7EF, UK

Address correspondence to Delia Fuhrmann, MRC Cognition and Brain Science Unit, 15 Chaucer Road, Cambridge, CB2 7EF, UK.
Email: delia.fuhrmann@mrc-cbu.cam.ac.uk.

Abstract

Fluid intelligence is the capacity to solve novel problems in the absence of task-specific knowledge and is highly predictive of outcomes like educational attainment and psychopathology. Here, we modeled the neurocognitive architecture of fluid intelligence in two cohorts: the Centre for Attention, Leaning and Memory sample (CALM) ($N = 551$, aged 5–17 years) and the Enhanced Nathan Kline Institute—Rockland Sample (NKI-RS) ($N = 335$, aged 6–17 years). We used multivariate structural equation modeling to test a preregistered watershed model of fluid intelligence. This model predicts that white matter contributes to intermediate cognitive phenotypes, like working memory and processing speed, which, in turn, contribute to fluid intelligence. We found that this model performed well for both samples and explained large amounts of variance in fluid intelligence ($R^2_{CALM} = 51.2\%$, $R^2_{NKI\text{-}RS} = 78.3\%$). The relationship between cognitive abilities and white matter differed with age, showing a dip in strength around ages 7–12 years. This age effect may reflect a reorganization of the neurocognitive architecture around pre- and early puberty. Overall, these findings highlight that intelligence is part of a complex hierarchical system of partially independent effects.

Key words: fractional anisotropy, processing speed, structural equation modeling, watershed model, working memory

Introduction

Fluid intelligence (g_f) is a core part of human cognition and refers to the capacity to solve novel problems in the absence of task-specific knowledge. It is highly predictive of a number of important life span outcomes, including educational attainment (Primi et al. 2010; Roth et al. 2015) and psychopathology (Gale et al. 2010). Despite years of investigation, our understanding of the neurocognitive architecture of g_f remains limited. Long-standing debates have considered, for instance, how g_f relates to more fundamental cognitive functions such as working memory and processing speed and how all of these cognitive functions relate to brain structure and function (Kyllonen and Christal 1990; Fry and Hale 2000; Chuderski 2013; Ferrer et al. 2013).

Working memory is the ability to hold and manipulate information in the mind short-term. It has been suggested that working memory is a key determinant of g_f by limiting mental information processing capacity (Fukuda et al. 2010; Chuderski 2013). Proponents of this working memory account of g_f cite high correlations between the two domains ranging from 0.5 to 0.9 in meta-analyses (Ackerman et al. 2005; Oberauer et al. 2005). Such high correlations have led some to suggest that g_f and working memory are isomorphic (Kyllonen and Christal 1990). However, more recent work has highlighted that this isomorphism only arises under conditions of high time constraints for g_f tasks (Chuderski 2013). This suggests that g_f and working memory are, in fact, separable constructs and underlines the importance of processing speed for g_f.

Processing speed, the speed of mental computations, is thought to be rate-limiting to g_f and is therefore sometimes proposed to be a particularly good predictor of g_f (Kail and Salthouse 1994; Salthouse 1996; Ferrer et al. 2013; Kail et al. 2015; Schubert et al. 2017). Proponents of the processing speed account of g_f cite moderate but robust correlations between g_f and processing speed of 0.2 in meta-analyses (Sheppard and Vernon 2008) as well as longitudinal evidence (Finkel et al. 2005; Coyle et al. 2011; Kail et al. 2015). Salthouse (1996) argued in the context of cognitive aging that processing speed is rate-limiting for high-level cognitive performance because slow processing means that relevant sub-operations cannot be completed in a given amount of time, or are not available for successful integration. A complementary explanation of individual differences in g_f proposes that processing speed may be a direct reflection of fundamental neuroarchitectonic properties of the brain, such as myelination or white matter microstructure (Lu et al. 2011; Chevalier et al. 2015).

White matter shows protracted development throughout childhood and adolescence and into the third decade of life (Mills et al. 2016). White matter tracts can be characterized *in vivo* using diffusion tensor imaging (DTI), which is sensitive, but not necessarily specific, to white matter microstructural properties such as myelination or axonal density (Jones et al. 2013; Wandell 2016). Fractional anisotropy (FA) is the most commonly investigated DTI measure and quantifies the directionality of water diffusion in different white matter tracts (Pfefferbaum et al. 2000; Wandell 2016). Working memory, processing speed, and g_f have each been linked to individual differences in FA (Vestergaard et al. 2011; Kievit et al. 2016; Bathelt et al. 2019). While some studies have posited that FA in different tracts can be summarized by sizable single components (Penke et al. 2010; Cox et al. 2016), formal investigations using confirmatory factor analysis have demonstrated that single-factor models of FA generally show poor fit and do not adequately capture individual differences in white matter microstructure (Lövdén et al. 2013; Kievit et al. 2016). In a similar vein, a growing body of literature has highlighted specific associations between white matter tracts and cognitive abilities, with those connecting frontoparietal regions usually showing largest contributions to complex cognitive functions like g_f (Vestergaard et al. 2011; Kievit et al. 2016; Bathelt et al. 2019).

We here seek to address several critical outstanding issues in the field: First, there is limited systematic evidence on the concurrent relationships between g_f, working memory, processing speed, and white matter. This leaves the relative contributions of processing speed and working memory to g_f unclear, which, in turn, poses challenges for the design of effective cognitive training interventions. Second, studies usually use a single task as a proxy for complex and abstract constructs such as processing speed, working memory, and g_f. This raises questions about the generalizability of findings (Noack et al. 2014). Third, our understanding of how the relationships between relevant cognitive domains and between brain and cognition change with age remains limited (Garrett 1946; Johnson 2000; Tamnes et al. 2017).

To address these issues, we here used structural equation modeling (SEM) to model the associations between g_f, working memory, processing speed, and white matter microstructure and age in two large, independent samples: the Centre for Attention, Leaning and Memory sample (CALM, $N = 551$, aged 5–17 years), which consists of children and adolescents referred
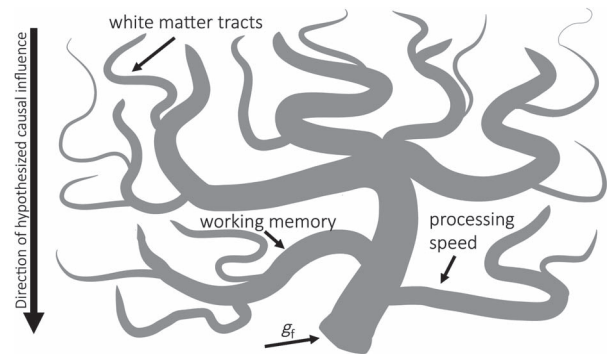

Figure 1. The watershed model. Schematic representation of the watershed model developed by Cannon and Keller (2006) and adapted for the present study. Fluid ability is hypothesized to be the downstream product of working memory and processing speed, which are, in turn, the product of white matter contributions. Figure adapted from Kievit et al. (2016).

to a clinic for having problems with attention, learning, and memory (Holmes et al. 2019); and the Enhanced Nathan Kline Institute—Rockland Sample (NKI-RS, $N = 335$, aged 6–17 years), a community-ascertained sample (Nooner et al. 2012).

To investigate the neurocognitive architecture of g_f in a principled way, we used a watershed model of individual differences. Based on the metaphor of a watershed, the model predicts a hierarchical many-to-one mapping of partially independent effects such that upstream tributaries (e.g., brain structure) contribute to intermediate cognitive phenomena (cognitive endophenotypes, e.g., working memory and processing speed), which then contribute to downstream, complex cognitive phenomena such as g_f (Cannon and Keller 2006; Kievit et al. 2016). See Figure 1 for a representation of the model.

SEM, as a statistical technique, is uniquely suited to model the complex multivariate brain–behavior associations posited by the watershed model (Kievit et al. 2011; Kline 2015). SEM combines factor analysis and path analysis (a variant of regression analysis). It can model abstract cognitive constructs like g_f, by estimating latent variables from observed task scores (i.e., manifest variables). This feature of SEM allowed us to model g_f, working memory, and processing speed in two independent samples and thereby provided a direct test of the generalizability of our findings. Second, SEM can test the simultaneous relations between multiple cognitive and neural variables, allowing us to address the relative contributions of different white matter tracts and different cognitive endophenotypes to g_f. Finally, using SEM Trees (Brandmaier et al. 2013), a novel decision-tree–based extension of SEM, we investigated whether the associations in the watershed model change with age.

Based on the watershed model, we made the following preregistered predictions (http://aspredicted.org/blind.php?x=u5pf6z):

1. Working memory, g_f, and processing speed are separable constructs.
2. Individual differences in g_f are predicted by working memory and processing speed.
3. White matter microstructure is a multidimensional construct.
4. There is a hierarchical relationship between white matter microstructure, cognitive endophenotypes (working memory and processing speed), and g_f, such that white matter

contributes to working memory and processing speed, which, in turn, contribute to g_f.

5. The contribution of working memory and processing speed to g_f change with age.

Materials and Methods

Samples

We analyzed data from the CALM and NKI-RS sample, as described in detail by Holmes et al. (2019) and Nooner et al. (2012), respectively. See also Simpson-Kent et al. (2019). We had also preregistered to analyze data from the Adolescent Brain Cognitive Development (ABCD) cohort (Volkow et al. 2018). The latter cohort contains only data for 9- and 10-year olds at present, however, which limits comparability to CALM and NKI-RS and currently makes it unsuitable for investigations of developmental differences. We therefore opted to not analyze ABCD data here and instead recommend a replication of the analyses presented here once longitudinal ABCD data are available. The CALM sample consists of children and adolescents referred by health and educational professionals as having difficulties in attention, learning, and/or memory. The NKI-RS is a community-ascertained, lifespan sample, and representative of the general population of Rockland, New York and the United States as a whole, in terms of ethnicity, socioeconomic status, etc. For NKI-RS, we included data for participants under the age of 18 only to match the age range of CALM and excluded data that were completed more than half a year after enrollment. The latter criterion was implemented to ensure that age at assessment did not differ substantively between cognitive measures. The final samples included 551 participants from CALM (30.85% female, aged 5.17–17.92 years, $N_{Neuroimaging} = 165$) and 335 participants from NKI-RS (44.48% female, aged 6.06–17.92 years, $N_{Neuroimaging} = 67$). See Table 1 for prevalence of relevant disorders and learning difficulties in the samples.

Cognitive Tasks

We included cognitive tasks measuring the domains of g_f, working memory, or processing speed for CALM and NKI-RS. See Table 2 for the complete list of tasks used and the Supplementary Methods for task descriptions. Supplementary Figures 1 and 2 show raw scores on all tasks. The tasks modeled here were preregistered for CALM but not NKI-RS.

Table 1 Prevalence of relevant disorders and learning difficulties in the CALM and NKI-RS cohorts

Variable	Percentage CALM	Percentage NKI-RS
ADHD	31.94	17.01
Dyslexia	5.81	5.67
Autism	6.72	0.60
Mood disorder	0.54	0.90
Anxiety disorder	2.36	18.21
Medicated[1]	10.53	17.01
Speech/language problems	38.11	19.40

Note. [1] unspecified medication for NKI-RS, ADHD medication for CALM. Diagnoses are not mutually exclusive.

White Matter Microstructure

We modeled mean FA for all 10 tracts of the Johns Hopkins University white matter tractography atlas (Hua et al. 2008) averaged over the hemispheres (Fig. 2). See Supplementary Methods for details of the magnetic resonance imaging (MRI) acquisition and processing and Supplementary Figures 3 and 4 for raw FA values in all tracts.

Analysis Methods and SEM

Covariance matrices and scripts replicating key analyses can be obtained from https://osf.io/4dusp/. Supplementary Figures 5 and 6 show correlation matrices of all tasks and white matter tracts modeled. We modeled raw scores for g_f and working memory tasks, as preregistered. Raw scores on processing speed tasks were transformed. This step was not preregistered but found necessary to achieve model convergence and to ensure interpretability of scores. First, we inverted response time scores (using the formula $y = 1/x$) to obtain more intuitive measures of "speed" for all but the Penn's Computerized Neurocognitive Battery (CNB) Motor Speed task, for which raw scores were already a measure of speed. Afterwards, we applied a log-transformation to reaction time tasks to increase normality and aid estimation. For the CNB Motor Speed task only, we additionally removed values ±2 standard deviation of the mean ($N = 6$) because the presence of these outliers had caused convergence problems.

We modeled the associations between cognition and white matter microstructure using SEM in R (R Core Team 2015) and the lavaan package (Rosseel 2012). All models were fit using maximum likelihood estimation with robust Huber–White standard errors and a scaled test statistic. Missing data were addressed using full information maximum likelihood estimation.

We used SEM Trees to investigate whether the associations among cognitive and neural measures differed with age. SEM Trees use decision-tree methods to hierarchically split a data set into subgroups based on a covariate of interest—in this case, age (Brandmaier et al. 2013). We first ran the watershed model in OpenMx (Boker et al. 2011) and then passed the model to semtree to compute the SEM Trees. We ran one SEM Tree for each parameter of interest (e.g., the covariance between working memory and processing speed). All other parameters in each semtree object were set to be invariant across groups to ensure that splits were specific to the parameter of interest. We used a 10-fold cross-validation estimation method as recommend by Brandmaier et al. (2013). For the path from the cingulate gyrus to working memory only, we used 5-fold cross-validation because the model did not converge using 10-fold cross-validation. Minimum sample size in age group was set to $N = 50$ to ensure reliable estimation of standard errors. Note that this choice effectively limited search space for potential splits to 6.58–12.42 years for CALM and 8.12–15.49 years for NKI-RS.

Results

To evaluate the hypotheses generated by the watershed model, we built up the watershed model in steps and carried our comprehensive tests of model fit at each step. First, we assessed the overall fit of our models to the data using the chi-square test, root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residual (SRMR). Good absolute fit was defined as RMSEA < 0.05, CFI > 0.97, and SRMR < 0.05 and acceptable

Table 2 Cognitive tasks modeled

	CALM	NKI-RS
Working memory	AWMA Digit Recall (Forward Digit Span)	WISC-R Forward Digit Span
	AWMA Backward Digit Span	WISC-R Backward Digit Span
	AWMA Dot Matrix	—
	AWMA Mr X	—
	—	CNB N-Back Task
g_f	WASI-II Matrix Reasoning	WASI-II Matrix Reasoning
	—	WASI-II Block Design
	—	WASI-II Similarities
	—	CNB Verbal Reasoning
Processing speed	DKEFS Trail-Making	DKEFS Trail-Making
	PhAB Rapid Naming	—
	TEA-Ch RBBS	—
	—	CNB Motor Speed
	—	CNB Sensory Motor Speed

Note. See the Supplementary Methods for task descriptions. Abbreviations: AWMA—Automated Working Memory Assessment (Alloway 2007), CNB—Computerized Neurocognitive Battery (Gur et al. 2001), DKEF—Delis–Kaplan Executive Functioning System (Delis et al. 2004), PhAB—Phonological Assessment Battery (Gallagher and Frederickson 1995), TEA-Ch RBBS—Test of Everyday Attention for Children, Red & Blues, Bags & Shoes subscale (Manly et al. 2001), WASI—Wechsler Abbreviated Scale of Intelligence—Second Edition (Wechsler 2011), WISC-R—Wechsler Intelligence Scale for Children—Revised (Kaufman 1975).
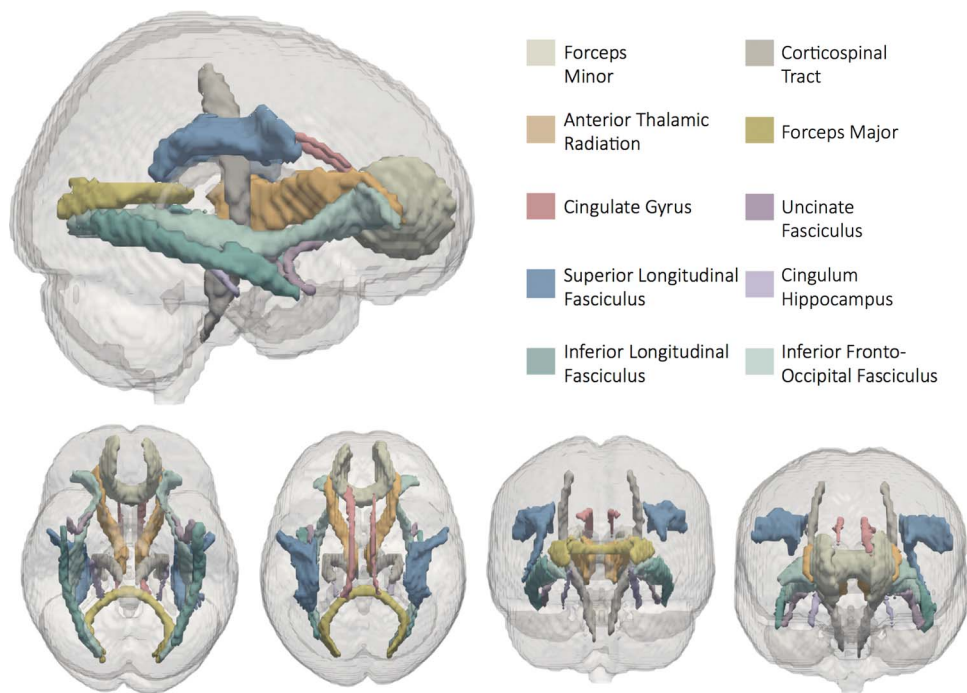

Figure 2. White matter tracts modeled in the analyses.

fit as RMSEA = 0.08–0.05, CFI = 0.95–0.97, and SRMR = 0.05–0.10 (Schermelleh-Engel et al. 2003). Second, we assessed specific predictions from our models by comparing them to alternative models. Comparative model fit for nested models was assessed using the chi-square difference test. Non-nested models were compared using the Akaike information criterion (AIC) weights, which indicates the probability of a model being the data-generating model compared with all other models tested (Wagenmakers and Farrell 2004). Lastly, we evaluated the significance and strength of relationships between specific variables in our models by inspecting the Wald test for individual parameters, noting the joint R^2 where relevant and reporting standardized parameter estimates. Absolute standardized parameter estimates above 0.10 were defined as small effects, 0.20 as typical, and 0.30 as large (Gignac and Szodorai 2016).

The Measurement Model of Cognition

To examine the neurocognitive architecture of g_f, we started by modeling the cognitive components of the watershed model: g_f, working memory, and processing speed. Specifically, we fit a 3-factor model of cognition (Fig. 3) and compared it with alternative measurement models. This approach allowed us to test Hypothesis 1, namely that g_f, working memory, and processing speed form three separable, albeit likely correlated cognitive factors.
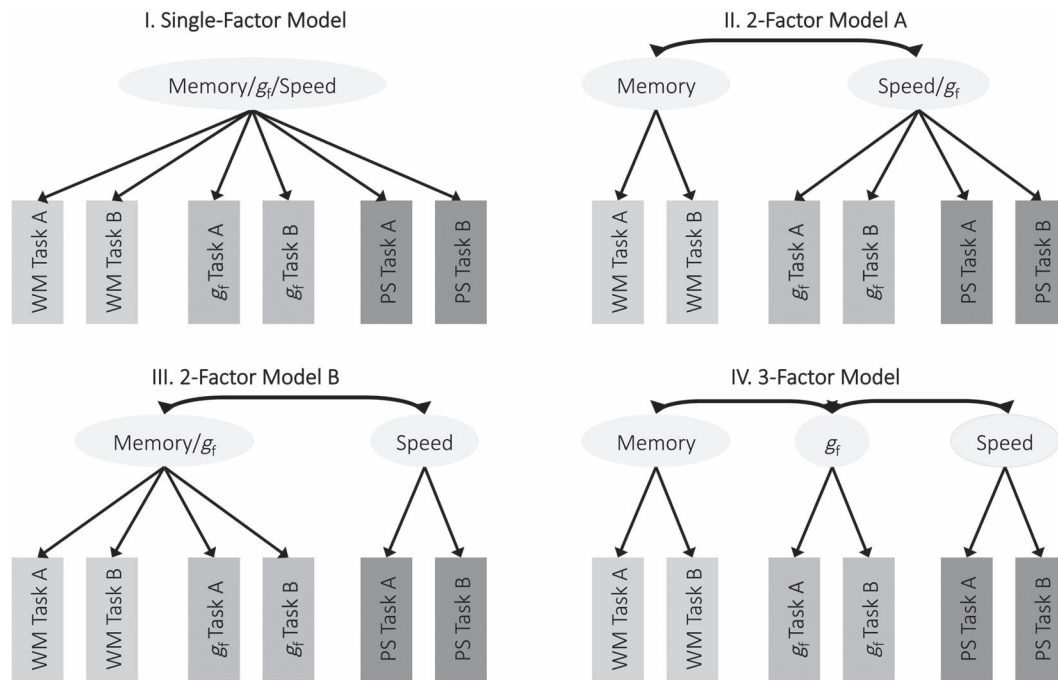
Figure 3. Different measurement models of cognition. Abbreviations: WM: working memory, PS: processing speed

The 3-factor model (Figure 3) showed excellent absolute fit for both the CALM and NKI-RS sample (Table 3), indicating that overall, the data were compatible with a model of g_f, working memory, and processing speed as 3 separate factors.

The 3-factor model also showed very good comparative fit for NKI-RS as well, with a 96.60% probability of being the data-generating model compared with all alternative models tested, as indicated by its AIC weight (Fig. 3). The evidence was more mixed for CALM, for which the 3-factor model showed a 27.15% probability of being the data-generating model, while 2-factor model B (Fig. 3, treating working memory and g_f as a unitary factor) showed a 72.85% probability of being the data-generating model, highlighting a close relationship between g_f and working memory for this sample. The single-factor model and 2-factor model A (Fig. 3, treating speed and g_f as a unitary factor) showed a very low (approximately 0%) probability of being the data-generating model, indicating that speed and g_f were clearly separable in both samples.

Overall, these results provide mixed evidence for Hypothesis 1: even though working memory, processing speed, and g_f were highly correlated (Table 4), processing speed formed a clearly separable factor from working memory and g_f in both samples. Working memory and g_f, however, were clearly separable only in NKI-RS but not CALM, suggesting greater similarity between g_f and working memory in the CALM sample. To facilitate comparison across samples and in accordance with our preregistered analysis plan, we nonetheless used the 3-factor measurement model (Table 4, Supplementary Table 1) in all subsequent analyses.

The Relationship Between Working Memory, Processing Speed, and g_f

We next examined the relationships between working memory, processing speed, and g_f in more detail. Specifically, we fit a SEM including regression paths between working memory and g_f, as well as speed and g_f, to test Hypothesis 2—that working memory and processing speed each predict individual differences in g_f. We found that this model showed good absolute fit for both samples (CALM: $\chi^2(18) = 41.74$, $P = 0.001$; RMSEA = 0.049 [0.030–0.068]; CFI = 0.983; SRMR = 0.032, NKI-RS: $\chi^2(32) = 54.15$, $P = 0.009$; RMSEA = 0.045 [0.024–0.065]; CFI = 0.981; SRMR = 0.030), indicating that, overall, the data were compatible with our model.

To further scrutinize the relationship between g_f, working memory, and speed, we compared our freely estimated model with a set of alternative models with different constraints imposed upon the regression paths. First, to test whether working memory and speed each made different contributions, we tested an alternative model in which the paths from processing speed and working memory to g_f were constrained to be equal. In CALM ($\Delta\chi^2(1) = 15.53$, $P < 0.001$), but not NKI-RS ($\Delta\chi^2(1) = 3.25$, $P = 0.072$), the freely estimated model fit better than the equality-constrained model, indicating that working memory and speed each made different contributions in CALM but not NKI-RS. Next, we tested whether the freely estimated model fit better than a model in which the path between g_f and working memory was constrained to zero. We found that that the freely estimated model fit better for both samples (CALM: $\Delta\chi^2(1) = 20.77$, $P < 0.001$; NKI-RS: $\Delta\chi^2(1) = 12.97$, $P < 0.001$). In line with our hypothesis, this result indicates that working memory makes a significant incremental contribution to g_f. Finally, we tested a model in which the path between g_f and processing speed was constrained to zero. This model showed no difference in fit to the freely estimated model for CALM ($\Delta\chi^2(1) = 0.02$, $P = 0.875$) or NKI-RS ($\Delta\chi^2(1) = 0.04$, $P = 0.849$). Contrary to our hypothesis, this indicates that there was no clear incremental contribution of processing speed to g_f.

Finally, we inspected standardized path estimates of the freely estimated model to assess the effect seizes of working memory and processing speed. Parameter estimates showed

Table 3 Model fit of competing measurement models

Single-factor model	2-factor model A	2-factor model B	3-factor model
		CALM	
$\chi^2(20) = 70.28, P < 0.001$	$\chi^2(19) = 67.99, P < 0.001$	$\chi^2(19) = 41.66, P = 0.002$	$\chi^2(18) = 41.74, P = 0.001$
RMSEA = 0.068	RMSEA = 0.068	RMSEA = 0.047	RMSEA = 0.049
[0.051–0.085]	[0.052–0.086]	[0.027–0.066]	[.030–.068]
CFI = 0.963	CFI = 0.964	CFI = 0.983	CFI = 0.983
SRMR = 0.047	SRMR = 0.043	SRMR = 0.032	SRMR = 0.032
AIC = 9697.18	AIC = 9696.44	AIC = 9668.58	AIC = 9670.55
BIC = 9800.66	BIC = 9804.24	BIC = 9776.37	BIC = 9782.66
$AIC_{weight} = 0\%$	$AIC_{weight} = 0\%$	$AIC_{weight} = 72.85\%$	$AIC_{weight} = 27.15\%$
		NKI-RS	
$\chi^2(35) = 109.96, P < 0.001$	$\chi^2(34) = 108.15, P < 0.001$	$\chi^2(34) = 64.85, P = 0.001$	$\chi^2(32) = 54.15, P = 0.009$
RMSEA = 0.080	RMSEA = 0.081	RMSEA = 0.052	RMSEA = 0.045
[0.064–0.097]	[0.064–0.098]	[0.033–0.071]	[0.024–0.065]
CFI = 0.936	CFI = 0.936	CFI = 0.974	CFI = 0.981
SRMR = 0.045	SRMR = 0.044	SRMR = 0.035	SRMR = 0.030
AIC = 7155.64	AIC = 7155.74	AIC = 7109.43	AIC = 7102.74
BIC = 7270.07	BIC = 7273.98	BIC = 7227.67	BIC = 7228.60
$AIC_{weight} = 0\%$	$AIC_{weight} = 0\%$	$AIC_{weight} = 3.40\%$	$AIC_{weight} = 96.60\%$

Note. See Figure 3 for the configuration of different models. Abbreviations: Akaike information criterion—AIC, Bayesian Information Criterion—BIC, Akaike weight—AIC_{weight}.

Table 4 Covariance between cognitive measures in the 3-factor model

Sample	Path	Standardized estimate
CALM	$g_f \leftrightarrow$ memory	0.71, z = 28.42, P < 0.001
	$g_f \leftrightarrow$ speed	0.55, z = 12.20, P < 0.001
	memory $\leftrightarrow$ speed	0.79, z = 19.35, P < 0.001
NKI-RS	$g_f \leftrightarrow$ memory	0.91, z = 19.51, P < 0.001
	$g_f \leftrightarrow$ speed	0.81, z = 24.73, P < 0.001
	memory $\leftrightarrow$ speed	0.87, z = 17.43, P < 0.001

Note. See Supplementary Table 1 for factor loadings.

that working memory showed a greater effect on g_f than processing speed, particularly in CALM (Table 5) even though raw correlations between g_f and speed were high in both samples (Table 4).

Overall, these results provide mixed evidence for Hypothesis 2: there was good evidence that working memory and speed made a significant joint contribution to g_f and that working memory made an incremental contribution to g_f in CALM. Contrary to our hypothesis, and the watershed model, however, processing speed showed no significant incremental contribution to g_f above and beyond working memory. We explore likely explanations for this finding in the Discussion.

The Measurement Model of White Matter

We next examined the measurement model of white matter to test Hypothesis 3, namely that white matter microstructure is a multidimensional construct. Specifically, we examined absolute model fit of a single-factor model to test whether a unidimensional model could adequately capture white matter microstructure. As expected, the single-factor model of white matter microstructure did not fit the data well (CALM: $\chi^2(35) = 124.63, P < 0.001$; RMSEA = 0.125 [0.103–.147]; CFI = 0.933;

SRMR = 0.039; NKI-RS: $\chi^2(35) = 132.33, P < 0.001$; RMSEA = 0.204 [0.167–0.242]; CFI = 0.885; SRMR = 0.023). This indicates that white matter microstructure could not be reduced to a single "global FA" dimension in our samples, in line with Lövdén et al. (2013), Kievit et al. (2016), and supporting Hypothesis 3. We therefore modeled each of the 10 white matter tracts separately in all subsequent models.

The Watershed Model: Relationships Between Cognition and White Matter

Next, we fit the full watershed model including white matter, working memory, processing speed, and g_f. Following our general analysis procedure, we investigated overall model fit, alternative models, and individual path estimates to gain a comprehensive understanding of the relationships in the watershed model and to test Hypothesis 4—that white matter contributes to working memory capacity and processing speed, which, in turn, contribute to g_f.

We found largely converging results across samples. The watershed model showed good absolute fit in CALM ($\chi^2(78) = 107.78, P = 0.014$; RMSEA = 0.026 [0.012–0.038]; CFI = 0.981; SRMR = 0.043) and acceptable fit in NKI-RS ($\chi^2(112) = 219.22$, $P < 0.001$; RMSEA = 0.053 [0.043–0.064]; CFI = 0.928; SRMR = 0.088). White matter explained large amounts of variance in working memory ($R^2_{CALM} = 32.3\%$; $R^2_{NKI-RS} = 46.1\%$) and processing speed ($R^2_{CALM} = 38.2\%$; $R^2_{NKI-RS} = 54.4\%$), which, in turn, explained even more variance in g_f ($R^2_{CALM} = 51.2\%$; $R^2_{NKI-RS} = 78.3\%$). In line with Hypothesis 4, this indicates that the watershed model fit the data overall.

Comparing the freely estimated watershed model to alternative, constrained, models showed that white matter contributed significantly to memory and processing speed. Specifically, a model in which paths from white matter to processing speed were constrained to zero fit worse than the freely estimated model (CALM: $\Delta\chi^2(10) = 50.26, P < 0.001$; NKI-

Table 5 Regression path estimates

Sample	Path	Standardized Estimate
CALM	speed → g_f	−0.01, z = −0.16, P = 0.876
	memory → g_f	0.72, z = 7.65, P < 0.001
NKI-RS	speed → g_f	0.06, z = 0.21, P = 0.836
	memory → g_f	0.86, z = 1.81, P = 0.070

RS: $\Delta\chi^2(10) = 27.19$, $P = 0.002$), as did a model in which paths from white matter to working memory were constrained to zero (CALM: $\Delta\chi^2(10) = 52.15$, p < 0.001; NKI-RS: $\Delta\chi^2(10) = 25.85$, $p = 0.004$). As hypothesized, white matter therefore contributed to both processing speed and working memory.

We next inspected the relationship between individual white matter tracts and working memory and speed in more detail. A model in which paths from white matter to working memory and speed were constrained to be equal fit worse than the freely estimated watershed model for CALM ($\Delta\chi^2(18) = 47.76$, P < 0.001) and NKI-RS ($\Delta\chi^2(18) = 30.42$, P = 0.034), indicating that the role of white matter microstructure in supporting working memory and processing speed differed across tracts. This supports the notion that there is a many-to-one mapping between white matter and cognition—a core tenet of the watershed model.

Investigating individual standardized parameter estimates of the different white matter tracts showed that for CALM, only the anterior thalamic radiation contributed significantly to processing speed, whereas the superior longitudinal fasciculus, forceps major, and cingulate gyrus were significantly, independently, and positively related to working memory (Fig. 4). For NKI-RS, the superior longitudinal fasciculus was significantly and positively related to processing speed and working memory

(Fig. 5). Two tracts showed an unexpected, strongly negative (<−1), relationship: the forceps minor for CALM and the inferior fronto-occipital fasciculus for NKI-RS. We found that these negative estimates occurred only when all other brain to cognition pathways were also estimated; when estimated on their own, path estimates were positive (forceps minor to working memory: standardized estimate = 0.36, z = 4.05, P < 0.001; inferior fronto-occipital fasciculus to working memory: standardized estimate = 0.14, z = 0.86, P = 0.390; inferior fronto-occipital fasciculus to processing speed: standardized estimate = 0.26, z = 1.41, P = 0.158). This sign-flip suggests that the negative pathways were potentially due to modeling several highly correlated paths at the same time (Jöreskog 1999). Overall, these results further support the watershed prediction that multiple white matter tracts map onto working memory and processing speed.

Finally, we probed the watershed model in more detail by testing a set of alternative expressions of the watershed model still compatible with the core tenets of the watershed model—as well as a set of alterative models incompatible with the watershed model. We compared all alternatives (see Fig. 6 for graphical representations) with the original watershed model by inspecting each model's relative probability of being the data-generating model as indicated by AIC weights (Wagenmakers and Farrell 2004). We found that the original watershed model showed a very high probability (98.58%) of being the data-generating model for CALM but only a 0.10% probability for NKI-RS. For NKI-RS, a different expression of the watershed model, such that g_f was regressed on working memory, which was regressed on processing speed, which was then regressed on white matter (Alternative A, Fig. 6), showed a 97.04% probability of being the data-generating model. This model only showed a 0.37% probability for CALM. Another expression of the watershed model, in which all tasks were
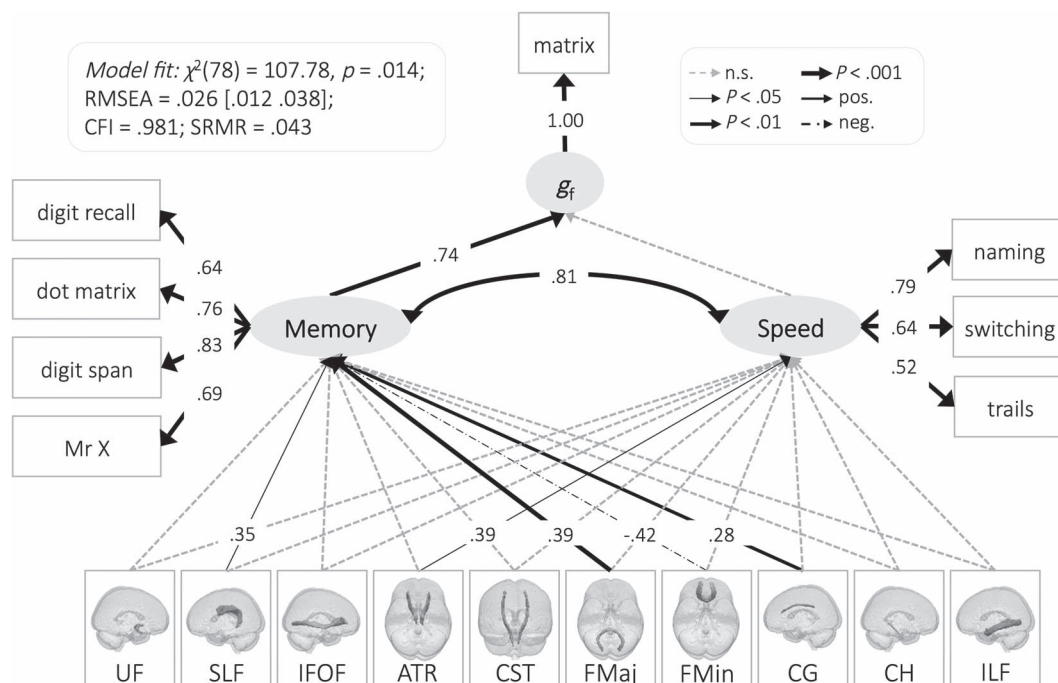

Figure 4. Watershed model in CALM. See Supplementary Table 2 for regression estimates. Residual covariances between white matter tracts were allowed but are not shown for simplicity. Abbreviations: uncinate fasciculus: UF, superior longitudinal fasciculus: SLF, inferior fronto-occipital fasciculus: IFOF, anterior thalamic radiations: ATR, cerebrospinal tract: CST, forceps major: FMaj, forceps minor: FMin, dorsal cingulate gyrus: CG, ventral cingulate gyrus: CH, inferior longitudinal fasciculus: ILF.
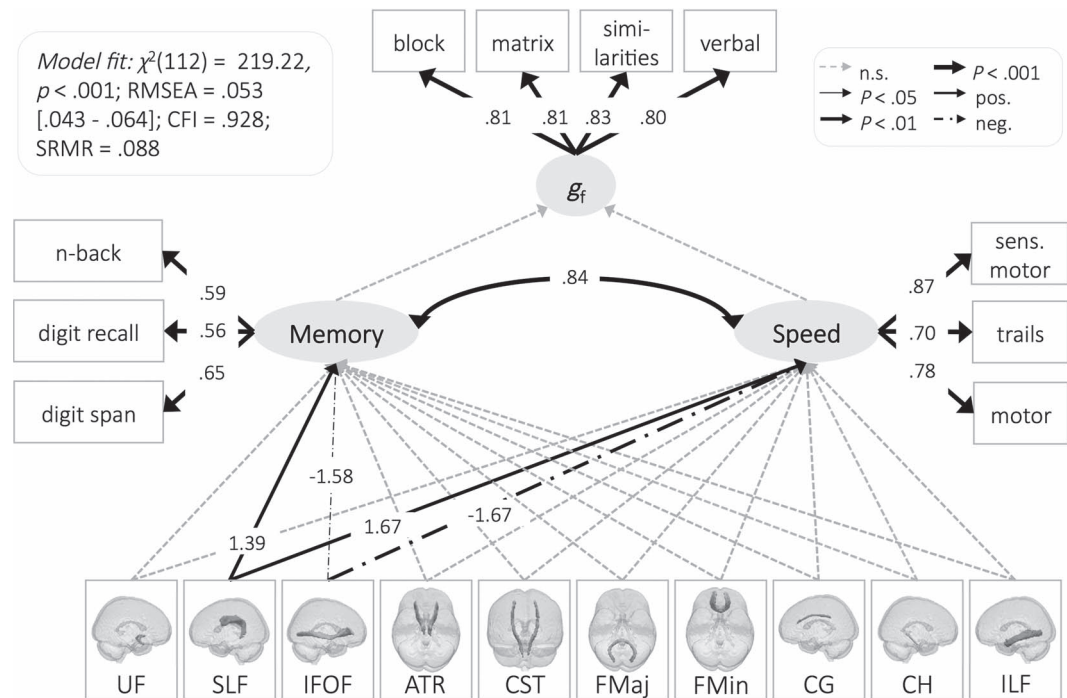
Figure 5. The watershed model in NKI-RS. See Supplementary Table 3 for regression estimates. Residual covariances between white matter tracts were allowed but are not shown for simplicity.
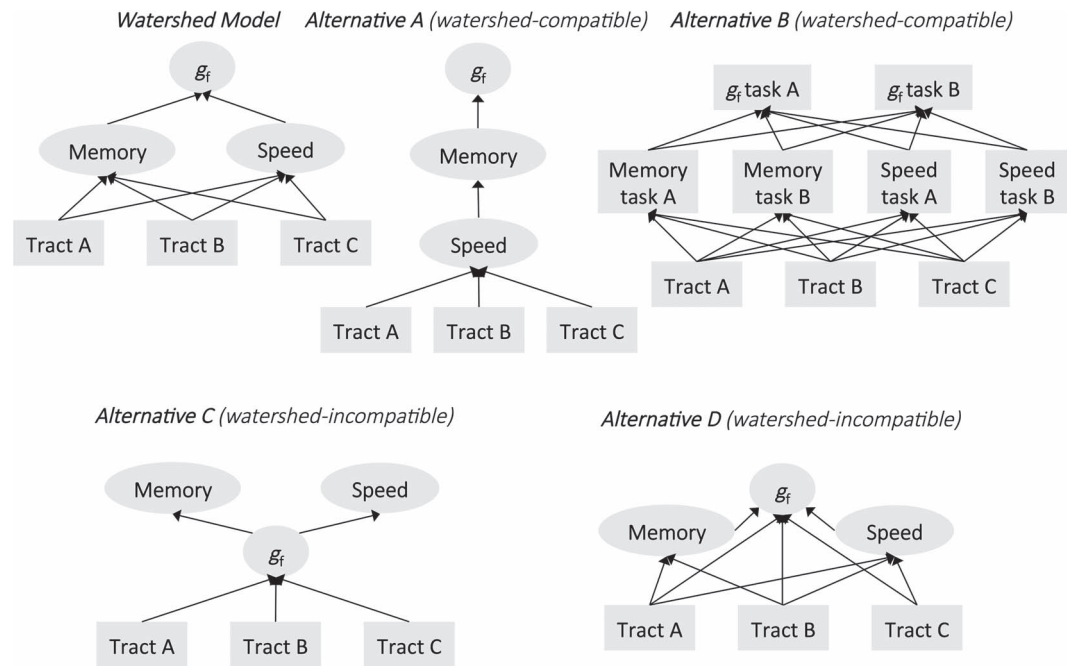

Figure 6. Configuration of alternative models. Alternatives A and B are watershed compatible, while C and D are watershed incompatible. The best-fitting model for CALM is the original watershed model; the best-fitting model for NKI-RS is Alternative A. Regression paths only are shown for simplicity. Square shapes denote manifest variables and oval shapes latent variables.

modeled separately as manifest, rather than latent, variables (Alternative B, Fig. 6), showed no advantage over the watershed model for CALM (0.00% probability) or NKI-RS (0.00% probability). We next tested 2 alternative models incompatible with the tenets of the watershed model. We found that a model in

which the hierarchy between cognitive endophenotypes and g_f was inverted (Alternative C, Fig. 6) showed comparatively low probability of being the data-generating model for both CALM (0.00%) and NKI-RS (2.86%). Similarly, a model in which g_f was directly regressed on white matter, working

memory, and processing speed (Alternative D, Fig. 6) showed no clear advantage over the watershed model for CALM (1.05% probability) or NKI-RS (0.00% probability). Overall, these model comparisons highlight that while the watershed model fit the data for both samples and had large explanatory power (as indicated by R^2s), the precise configuration of the watershed model may differ somewhat between cohorts.

In summary, we found that the watershed model performed well overall for both cohorts. As hypothesized, white matter contributed to working memory and processing speed, which, in turn, contributed to g_f, and explained large amounts of variance therein. Also as predicted by the watershed model, there was a many-to-one mapping between white matter tracts and cognition. The exact configuration of the watershed model, however, may differ slightly between cohorts. These differences may be a function of cohort differences in sample size, average levels of cognitive ability, and/or the specific tasks used—a topic we will return to in the Discussion.

Testing for Potential Confounds

We carried out a series of supplementary and non-preregistered analyses to examine whether possible confounders influenced our models. These analyses showed that our findings were robust to the inclusion of covariates such as scanner motion or socio-economic status. They were also robust across genders and participants taking or not taking medication. There were no differences in the structure of the model between participants with and without diagnosed disorders for CALM. Potential small differences cannot be ruled out for NKI-RS, for which the multi-group model showed poor fit. This is likely due to the low number of diagnosed participants of N = 106 (see Supplementary Analyses).

Age-Related Differences in the Neurocognitive Architecture of g_f

Finally, we tested Hypothesis 5—that the contribution of working memory and processing speed to g_f varied with age. We first inspected cross-sectional differences in g_f, working memory, and processing speed and then used SEM Trees to investigate potential age differences in the relationships between these factors. In additional non-preregistered analyses, we also used SEM Trees to investigate potential age differences in the relationship between white matter and cognitive endophenotypes by inspecting paths that were positive and significant in the watershed model (Figs 4 and 5).

SEM Trees combine SEMs with decision tree methods, separating a data set into subgroups (in this case age groups) if SEM parameter estimates of interest differ sufficiently (Brandmaier et al. 2013). SEM Trees allowed us to investigate age as a potential moderator without imposing a priori categorical age splits. We initially allowed for no more than 2 age groups. This yielded inconsistent results for CALM and NKI-RS (see Supplementary Table 4). To test whether these inconsistencies were an artifact of allowing for only 2 groups, we repeated our analysis and allowed for up to 4 age groups. This analysis yielded consistent results between CALM and NKI-RS (Table 6). This pattern of results indicates that the initial parameters of our analysis caused us to miss relevant age differences.

As shown in Figure 7, g_f, working memory, and processing speed factor scores increased with age for all 3 cognitive phenotypes. In line with our hypothesis, SEM Trees showed
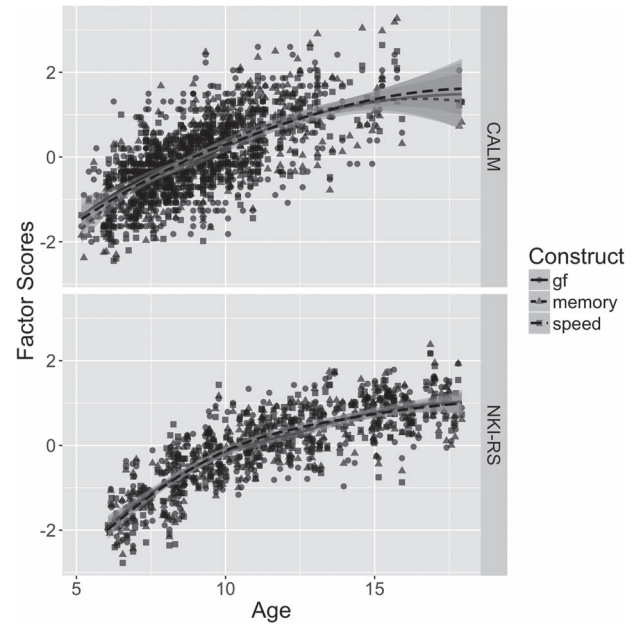

Figure 7. Cognitive factor scores by age.

that there were pronounced age-related differences in brain–behavior in childhood and adolescence (Table 6). For both samples and all but one path, there was an initially strong relationship between components of the watershed model, then a dip around ages 7–9 years for CALM and age 8 years for NKI-RS, followed by an increase in path strength around ages 11–12 years (see Supplementary Figure 7 for a graphical representation of these results). Speculatively, this pattern of results is consistent with an interpretation of a reorganization of neurocognitive faculties in late childhood, followed by a consolidation of neurocognitive pathways around the onset of adolescence (Johnson 2000, 2011).

Discussion

We here used multivariate statistical techniques to investigate the neurocognitive architecture of g_f in 2 large (N_{CALM} =551, N_{NKI-RS} =335) developmental cohorts and, for the first time, investigated how the neurocognitive architecture of g_f changes dynamically with age. We tested a preregistered watershed model of g_f, which predicts a hierarchy of partially independent effects. As might be expected from a multi-cohort study, there were some differences between the community-ascertained cohort (NKI-RS) and the cohort of children and adolescents with learning difficulties (CALM) in specific path estimates. Overall, however, we found convergent results across these 2 heterogeneous samples. The watershed model performed well for both CALM and NKI-RS; white matter contributed to working memory and processing speed, which, in turn, contributed to g_f and explained 51% of variance therein for the CALM sample and 78% of variance for NKI-RS. Models were robust across genders, participants taking or not taking medication, and when controlling for socio-economic status and scanner motion. Investigations of age effects showed that the relationship between cognitive abilities and white matter dipped in strength around ages 7–12 years. Speculatively, this age effect may reflect a reorganization of the neurocognitive architecture during pre-puberty and early puberty

Table 6 SEM Tree results for the watershed model

Path	Estimate Before	Age Split 1	Estimate Between	Age Split 2	Estimate Between	Age Split 3	Estimate After
CALM							
memory ↔ speed	0.85	**8.46**	0.97	**9.46**	0.74	—	—
memory → g_f	0.83	**9.38**	0.42	**10.04**	1.14	**10.88**	0.94
speed → g_f	0.04	**6.88**	−0.19	**11.21**	0.17	—	—
SLF → memory	0.67	**7.21**	0.18	**11.21**	0.76	—	—
FMaj → memory	0.59	**7.71**	0.14	**9.29**	0.33	**11.13**	0.74
CG → memory [1]	0.64	**6.96**	0.09	**11.04**	0.70	—	—
ATR → speed	0.96	**7.13**	0.68	**7.96**	0.17	**11.96**	0.65
NKI-RS							
memory ↔ speed	0.90	**9.82**	0.48	**14.72**	1.11	—	—
memory → g_f	1.10	**8.59**	0.59	**12.67**	1.03	—	—
speed → g_f	0.53	**8.59**	−0.12	**12.96**	0.52	—	—
SLF → memory	2.15	**8.30**	1.47	**12.15**	1.93	—	—
SLF → speed	3.12	**8.63**	1.83	**15.09**	2.31	—	—

Note. The table shows differences in parameter estimates for paths of interest (as shown in Figs 4 and 5) depending on participants' age in years. Our analyses allowed for a maximum of 3 age splits (and thus 4 age groups). An absence of a third age split (denoted by"—"in the table) indicates that the SEM tree split only twice, suggesting no further changes in parameter strength after the second split. See Supplementary Figure 7 for a graphical representation of these results.

(Byrne et al. 2017). These findings have implications for understanding and targeting cognitive impairments in populations with learning difficulties.

The watershed model tested here consists of 3 levels: g_f forms the most down-stream point, with working memory and processing speed as intermediate tributaries, and white matter microstructural tracts as upstream sources. Previous studies suggested that matter microstructure is best characterized by a single global FA factor (Penke et al. 2010), while others have contended that association patterns among different white matter tracts are more complex (Lövdén et al. 2013; Kievit et al. 2016). Here we found strong evidence for a multifactorial view of white matter tracts—for both samples, a unidimensional model of white matter fit poorly and for CALM, multiple tracts also showed partially independent contributions to distal cognitive outcomes. This is in line with the watershed model. There were some differences between cohorts as to which tracts contributed most to working memory and processing speed: In line with previous research (Kievit et al. 2016; MacPherson et al. 2017; Bathelt et al. 2019), we found that the anterior thalamic radiation was related to processing speed, as were the forceps major, forceps minor, and the cingulate gyrus to working memory for CALM. However, these tracts were not significant for NKI-RS. A possible explanation for these differences between samples is the discrepancy in the number of participants with imaging data (N = 165 in CALM vs. N = 67 in NKI-RS). This discrepancy likely confers differential power to detect weaker pathways. Other, not mutually exclusive, explanations are that the observed differences reflect differences in brain–behavior mapping between more atypical and typical cohorts (Bathelt et al. 2019), sampling variance across 2 independent cohorts collected under somewhat different socio-economic conditions (United Kingdom and United States of America), or a more uniform age distribution in NKI-RS. While DTI images were processed with the same pipeline across sites, the scanner and MRI acquisition protocol were also different. Although previous work suggests that FA is relatively robust measure in multi-site comparisons

(Vollmar et al. 2010), we cannot rule out site differences as a potential confound. It will be necessary to replicate these findings in large typical and atypical cohorts collected in the same setting. Of note, however, the superior longitudinal fasciculus was robustly associated with working memory across the 2 different samples and settings. For NKI-RS, the superior longitudinal fasciculus was also associated with processing speed. The superior longitudinal fasciculus is a large bilateral association fiber connecting temporal, occipital, parietal, and frontal regions (Kamali et al. 2014). It is therefore well situated for supporting cognitive processes such as g_f, which rely on integrative multiple-demand systems (Jung and Haier 2007; Fedorenko et al. 2013; Parlatini et al. 2017).

Our findings for the cognitive levels of the watershed model highlighted a close relationship between working memory and g_f. Previous studies had variably suggested that g_f and working memory (Kyllonen and Christal 1990; Fukuda et al. 2010) or g_f and processing speed (Kail and Salthouse 1994; Salthouse 1996; Coyle et al. 2011; Ferrer et al. 2013) may be most closely related. We found that all 3 cognitive factors were highly correlated for both samples. Nonetheless, processing speed formed a cognitive factor clearly separable from working memory and g_f. Working memory and g_f, in turn, were separable in the community-ascertained NKI-RS but not in CALM, the cohort of children and adolescents with learning difficulties. This close relationship between g_f and working memory was also evident in other models of CALM where processing speed and working memory were used as joint predictors of g_f: contrary to our hypotheses, processing speed became nonsignificant after controlling for working memory. There are several possible, and not mutually exclusive, explanations for this finding and the apparent differences between cohorts. First, a broader set of speed tasks (which might be captured by several latent variables for clerical speed, choice reaction time, and speed variability) might show higher predictive power than the single latent variable for speed, which could be modeled here. This may be particularly pertinent for heterogeneous cohorts like CALM. Second, and in line with previous work showing that time-constraints increase

isomorphism of g_f and working memory (Chuderski 2013), even standard implementations of g_f tasks may place considerable time pressure on struggling learners, thereby increasing g_f–working memory covariance in CALM as compared with NKI-RS. Conversely, less subjective or objective time pressure may also confer a differentiation of cognitive domains and the watershed hierarchy in cohorts of older ages and/or higher ability levels, such as NKI-RS. There was some evidence that, for this sample, speed formed an intermittent level in the hierarchy between white matter and working memory (Alternative A, Fig. 6). Future longitudinal research will be necessary to differentiate these alternative configurations of the watershed model and scrutinize the causal flow of effects. For now, our findings highlight the value of replicating analyses in different cohorts using different tasks. While evidence was mixed for the association between g_f and processing speed, the strong associations between g_f and working memory across samples indicate a robust and likely generalizable relationship between these two domains, supporting the notion that mental information processing capacity is a key determinant of g_f (Kyllonen and Christal 1990; Fukuda et al. 2010).

The associations in the watershed model differed between ages in a complex, non-monotonic fashion. Previous research suggested either a decrease in covariance among cognitive domains with age (age differentiation; Garrett 1946), an increase in covariance with age (age de-differentiation; Blum and Holling 2017), or no changes with age (Tucker-Drob 2009; de Mooij et al. 2018). These investigations have traditionally focused on relations between cognitive domains, however, not on relationships between brain and cognition—although see de Mooij et al. (2018). Possible linear and nonlinear changes in brain–behavior mapping with age have remained mostly unexplored (Tamnes et al. 2017). Using SEM Trees, a novel decision-tree–based technique, we here found evidence of complex developmental differences consistent across samples and relationships in the watershed model. Initially strong path estimates showed a pronounced decrease in strength around ages 7–9 years, followed by a renewed increase in the strength, even surpassing initial levels, around ages 10–15 years.

There are at least 2 possible explanations for this developmental dip in brain–cognition relationships. First, there may be a true decrease in relationship strength during this time of life. Possibly other cognitive skills such as verbal reasoning, temporarily support g_f, resulting in weaker relationships between g_f and working memory. Alternatively, the configuration of the watershed model may change temporarily during this time, which could also manifest in an apparently weaker covariance structure. In this case, the true relationship between g_f, memory, speed, and white matter may still be strong, just configured differently from the watershed model. We note that both explanations are compatible with the interactive specialization theory (Johnson 2000, 2011), which predicts as remapping of the relationships between brain substrates and cognitive abilities during development.

On a physiological level, this age effect may be driven by neuroendocrine changes during pre- and early puberty. Puberty is driven by a complex and only partially understood set of hormonal events including gonadarche and andrenarche (Sisk and Zehr 2005). Gonadarche begins with the secretion of gonadotropin-releasing hormone from the hypothalamus around ages 10–11 years and closely tracks the overt bodily changes of puberty (Dorn 2006). Andrenarche, beginning with the maturation of the andrenal gland, starts

as early as 6 years of age and is increasingly recognized as a complimentary driver of puberty and brain development (Byrne et al. 2017). It is possible that the hormonal changes of andrenarche and early gonadarche may lead to a level of neural reorganization, which may initially appear as weaker relationships in the watershed model. The sweeping bodily, social, and cognitive changes happening in early adolescence may then drive a consolidation of the neurocognitive architecture of g_f.

On a more general level, these age effects suggest the existence of potential nonlinear changes in brain–behavior mapping during childhood and adolescence and underline the value of modern statistical approaches, such as SEM Trees, for the study of age-related differences. It is worth noting, however, that these findings, which are based on an inherently exploratory technique, will need to be replicated in future confirmatory studies with fine-grained data on puberty and larger sample sizes. The latter will also allow for detailed investigations of potential gender differences. Moreover, while we were able to investigate individual differences in g_f, we could not assess intra-individual changes during childhood and adolescence. Although the relatively narrow age range makes large cohort effects unlikely, it may still be that there were differences in recruitment and selection that varied across the age range. As such, the cross-sectional nature of our samples limits our ability to make inferences about developmental dynamics.

Our study illustrates some of the advantages and challenges of preregistered secondary data analyses. We agree with others in the field that secondary data analysis need not be and should not be confounded with purely exploratory research (Mills and Tamnes 2014; Orben and Przybylski 2019; Scott and Kline 2019). Preregistrations, as well as dedicated multivariate methods such as SEM, can help reduce the scope for analytic flexibility and increase scientific rigor when using rich, secondary data sets. Preregistrations also do not preclude the use of exploratory methods or the ability to ask exploratory questions, as we did in our analysis of age effects. Preregistrations merely facilitate the distinction between exploratory and confirmatory research (Wagenmakers et al. 2012). There are, however, some unique challenges to preregistering secondary data analyses that are worth noting. First, information on the precise measures collected is not always available prior to data access, which can limit the level of detail in which an analysis can be preregistered. Second, data quality and the level of data processing, the latter being particularly relevant for MRI data, are not always clear a priori (e.g., see Kievit et al. 2018), which can necessitate changes to analyses plans after data inspection. Third, convergence issues are fairly common when using complex multivariate methods such as SEM. We found it necessary to transform some of our speed variables, for instance, to achieve model convergence. Such post hoc modifications, not guided by the palatability of the results, but rather by unforeseen, and sometimes unforeseeable, practical considerations, mean that preregistration can sometimes fall short of full compliance. Nevertheless, we believe that even imperfect preregistrations, alongside shared code, data, and the transparent presentation of results, can help the reader distinguish between confirmatory and exploratory results and adjust their level of confidence in conclusions accordingly. For guidance on maximizing transparency in preregistration of secondary data, see Weston et al. (2018).

Finally, the findings from our study have implications in understanding and targeting cognitive impairments in popu-

lations with learning difficulties. First, the close relationship between working memory and g_f found here and in other studies (Fukuda et al. 2010; Chuderski 2013) indicates that children and adolescents struggling with working memory are likely to also struggle in terms of complex reasoning tasks. Either reducing working memory load, decreasing time constraints, or training working memory and fluid ability capacity in such populations may therefore be promising lines of inquiry for intervention studies. It is worth highlighting, however, that cognitive training studies have so far shown little evidence of far transfer; training abstract reasoning, a common measure of g_f, has not resulted in robust increases in working memory (Knoll et al. 2016), and working memory training has not been shown to transfer to reasoning skills or school performance (Dunning et al. 2013; Schwaighofer et al. 2015). Similarly, transfer from processing speed to reasoning seems to be limited (Mackey et al. 2011). The results obtained here suggest that interventions may increase their chance of success by implementing programs of sufficient complexity to affect the entire neurocognitive architecture of effects (see also Kievit et al. 2016). The level of intensity required to produce sustained benefits may need to be as demanding and consistent as education itself, which shows robust effects in increasing general cognitive abilities over time (Ritchie and Tucker-Drob 2018). Our work and work by others (Noack et al. 2014) also highlight the value of assessing, modeling, and potentially intervening on multiple tasks, rather than relying on a single task to capture complex cognitive domains such as g_f. Finally, the age-related differences in the relationships of the watershed model observed using SEM Trees suggest that some interventions may work best at particular developmental phases.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Funding

Notes

Authors' Contributions

D.F. and R.A.K. conceived and designed the study. J.B. processed the imaging data. D.F., I.L.S.-K. and R.A.K. analyzed the data. All authors contributed to the write-up of the study.

References

Ackerman PL, Beier ME, Boyle MO. 2005. Working memory and intelligence: the same or different constructs? *Psychol Bull.* 131:30–60.

Alloway TP. 2007. *Automated working memory assessment.* London: The Psychological Corporation

Bathelt J, Johnson A, Zhang M, Astle DE. 2019. The cingulum as a marker of individual differences in neurocognitive development. *Scientific Reports.* 9:2281.

Blum D, Holling H. 2017. Spearman's law of diminishing returns. A meta-analysis. *Intelligence.* 65:60–66.

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T *et al.* 2011. OpenMx: an open source extended structural equation modeling framework. *Psychometrika.* 76:306–317.

Brandmaier AM, von Oertzen T, McArdle JJ, Lindenberger U. 2013. Structural equation model trees. *Psychol Methods.* 18:71–86.

Byrne ML, Whittle S, Vijayakumar N, Dennison M, Simmons JG, Allen NB. 2017. A systematic review of adrenarche as a sensitive period in neurobiological development and mental health. *Sensitive Periods Dev.* 25:12–28.

Cannon TD, Keller MC. 2006. Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol.* 2:267–290.

Chevalier N, Kurth S, Doucette MR, Wiseheart M, Deoni SCL, Dean DC III, O'Muircheartaigh J, Blackwell KA, Munakata Y, LeBourgeois MK. 2015. Myelination is associated with processing speed in early childhood: preliminary insights. *PLoS One.* 10:e0139897.

Chuderski A. 2013. When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence.* 41:244–262.

Cox SR, Ritchie SJ, Tucker-Drob EM, Liewald DC, Hagenaars SP, Davies G, Wardlaw JM, Gale CR, Bastin ME, Deary IJ. 2016. Ageing and brain white matter structure in 3513 UK Biobank participants. *Nat Commun.* 7:13629.

Coyle TR, Pillow DR, Snyder AC, Kochunov P. 2011. Processing speed mediates the development of general intelligence (g) in adolescence. *Psychol Sci.* 22:1265–1269.

de, Mooij SMM, Henson RNA, Waldorp LJ, Kievit RA. 2018. Age differentiation within grey matter, white matter and between memory and white matter in an adult lifespan cohort. *J Neurosci.* 38:5826–5836.

Delis DC, Kramer JH, Kaplan E, Holdnack J. 2004. Reliability and validity of the Delis-Kaplan executive function system: an update. *J Int Neuropsychol Soc.* 10:301–303.

Dorn LD. 2006. Measuring puberty. *J Adolesc Health.* 39:625–626.

Dunning DL, Holmes J, Gathercole SE. 2013. Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Dev Sci.* 16:915–925.

Fedorenko E, Duncan J, Kanwisher N. 2013. Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci U S A.* 110:16616–16621.

Ferrer E, Whitaker K, Steele J, Green C, Wendelken C, Bunge S. 2013. White matter maturation supports the development of

reasoning ability through its influence on processing speed. *Dev Sci.* 16:941–951.

Finkel D, Reynolds CA, McArdle JJ, Pedersen NL. 2005. The longitudinal relationship between processing speed and cognitive ability: genetic and environmental influences. *Behav Genet.* 35:535–549.

Fry AF, Hale S. 2000. Relationships among processing speed, working memory, and fluid intelligence in children. *Biol Psychol.* 54:1–34.

Fukuda K, Vogel E, Mayr U, Awh E. 2010. Quantity not quality: the relationship between fluid intelligence and working memory capacity. *Psychon Bull Rev.* 17:673–679.

Gale CR, Batty GD, Tynelius P, Deary IJ, Rasmussen F. 2010. Intelligence in early adulthood and subsequent hospitalisation and admission rates for the whole range of mental disorders: longitudinal study of 1 049 663 men. *Epidemiology.* 21:70–77.

Gallagher A, Frederickson N. 1995. The Phonological Assessment Battery (PhAB): an initial assessment of its theoretical and practical utility. *Educ Child Psychol.* 12:53–67.

Garrett HE. 1946. A developmental theory of intelligence. *Am Psychol.* 1:372–378.

Gignac GE, Szodorai ET. 2016. Effect size guidelines for individual differences researchers. *Pers Individ Dif.* 102:74–78.

Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE. 2001. Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology.* 25:766.

Holmes J, Bryant A, Gathercole SE. 2019. Protocol for a transdiagnostic study of children with problems of attention, learning and memory (CALM). *BMC Pediatr.* 19:1.

Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, Calabresi PA, Pekar JJ, van Zijl PCM, Mori S. 2008. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *NeuroImage.* 39:336–347.

Johnson MH. 2000. Functional brain development in infants: elements of an interactive specialization framework. *Child Dev.* 71:75–81.

Johnson MH. 2011. Interactive specialization: a domain-general framework for human functional brain development? *Dev Cogn Neurosci.* 1:7–21.

Jones DK, Knösche TR, Turner R. 2013. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *NeuroImage.* 73:239–254.

Jöreskog KG. 1999. How large can a standardized coefficient be? Retrieved April 10, 2018 from http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf

Jung RE, Haier RJ. 2007. The parieto-frontal integration theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav Brain Sci.* 30:135–154.

Kail RV, Lervåg A, Hulme C. 2015. Longitudinal evidence linking processing speed to the development of reasoning. *Dev Sci.* 19:1067–1074.

Kail RV, Salthouse TA. 1994. Processing speed as a mental capacity. *Acta Psychol (Amst).* 86:199–225.

Kamali A, Flanders AE, Brody J, Hunter JV, Hasan KM. 2014. Tracing superior longitudinal fasciculus connectivity in the human brain using high resolution diffusion tensor tractography. *Brain Struct Funct.* 219:269–281.

Kaufman AS. 1975. Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. *J Consult Clin Psychol.* 43:135–147.

Kievit R, Fuhrmann D, Borgeest G, Simpson-Kent I, Henson R. 2018. The neural determinants of age-related changes in fluid intelligence: a pre-registered, longitudinal analysis in UK Biobank. *Wellcome Open Res.* 3:38.

Kievit RA, Davis SW, Griffiths J, Correia MM, CamCAN, Henson RN. 2016. A watershed model of individual differences in fluid intelligence. *Neuropsychologia.* 91:186–198.

Kievit RA, Romeijn JW, Waldorp LJ, Wicherts JM, Scholte HS, Borsboom D. 2011. Mind the gap: a psychometric approach to the reduction problem. *Psychol Inq.* 22:67–87.

Kline RB. 2015. *Principles and practice of structural equation modeling.* New York, NY: Guilford Publications

Knoll LJ, Fuhrmann D, Sakhardande AL, Stamp F, Speekenbrink M, Blakemore SJ. 2016. A window of opportunity for cognitive training in adolescence. *Psychol Sci.* 27:1620–1631.

Kyllonen PC, Christal RE. 1990. Reasoning ability is (little more than) working-memory capacity?! *Intelligence.* 14:389–433.

Lövdén M, Laukka EJ, Rieckmann A, Kalpouzos G, Li T, Jonsson T, Wahlund L, Fratiglioni L, Bäckman L. 2013. The dimensionality of between-person differences in white matter microstructure in old age. *Hum Brain Mapp.* 34:1386–1398.

Lu PH, Lee GJ, Raven EP, Tingus K, Khoo T, Thompson PM, Bartzokis G. 2011. Age-related slowing in cognitive processing speed is associated with myelin integrity in a very healthy elderly sample. *J Clin Exp Neuropsychol.* 33:1059–1068.

Mackey AP, Hill SS, Stone SI, Bunge SA. 2011. Differential effects of reasoning and speed training in children. *Dev Sci.* 14:582–590.

MacPherson SE, Cox SR, Dickie DA, Karama S, Starr JM, Evans AC, Bastin ME, Wardlaw JM, Deary IJ. 2017. Processing speed and the relationship between Trail Making Test-B performance, cortical thinning and white matter microstructure in older adults. *Cortex.* 95:92–103.

Manly T, Anderson V, Nimmo-Smith I, Turner A, Watson P, Robertson IH. 2001. The differential assessment of children's attention: The Test of Everyday Attention for Children (TEA-Ch), normative sample and ADHD performance. *J Child Psychol Psychiatry.* 42:1065–1081.

Mills KL, Goddings A-L, Herting MM, Meuwese R, Blakemore S-J, Crone EA, Dahl RE, Güroğlu B, Raznahan A, Sowell ER *et al*. 2016. Structural brain development between childhood and adulthood: convergence across four longitudinal samples. *NeuroImage.* 141:273–281.

Mills KL, Tamnes CK. 2014. Methods and considerations for longitudinal structural brain imaging analysis across development. *Dev Cogn Neurosci.* 9:172–190.

Noack H, Lövdén M, Schmiedek F. 2014. On the validity and generality of transfer effects in cognitive training research. *Psychol Res.* 78:773–789.

Nooner K, Colcombe S, Tobe R, Mennes M, Benedict M, Moreno A, Panek L, Brown S, Zavitz S, Li Q *et al*. 2012. The NKI-Rockland Sample: a model for accelerating the pace of discovery science in psychiatry. *Front Neurosci.* 6:152.

Oberauer K, Schulze R, Wilhelm O, Süß H-M. 2005. Working memory and intelligence—their correlation and their relation: comment on Ackerman, Beier, and Boyle (2005). *Psychol Bull.* 131:61–65.

Orben A, Przybylski AK. 2019. The association between adolescent well-being and digital technology use. *Nat Hum Behav.* 3:173–182.

Parlatini V, Radua J, Dell'Acqua F, Leslie A, Simmons A, Murphy DG, Catani M, Thiebaut de Schotten M. 2017. Func-

tional segregation and integration within fronto-parietal networks. *NeuroImage*. 146:367–375.

Penke L, Maniega SM, Murray C, Gow AJ, Valdés Hernández MC, Clayden JD, Starr JM, Wardlaw JM, Bastin ME, Deary IJ. 2010. A general factor of brain white matter integrity predicts information processing speed in healthy older people. *J Neurosci*. 30:7569.

Pfefferbaum A, Sullivan EV, Hedehus M, Lim KO, Adalsteinsson E, Moseley M. 2000. Age-related decline in brain white matter anisotropy measured with spatially corrected echo-planar diffusion tensor imaging. *Magn Reson Med*. 44:259–268.

Primi R, Ferrão ME, Almeida LS. 2010. Fluid intelligence as a predictor of learning: a longitudinal multilevel approach applied to math. *Learn Individ Differ*. 20:446–451.

R Core Team, The. 2015. R: A language and environment for statistical computing.

Ritchie SJ, Tucker-Drob EM. 2018. How much does education improve intelligence? A meta-analysis. *Psychol Sci*. 29:1358–1369.

Rosseel Y. 2012. lavaan: an R package for structural equation modeling. *J Stat Softw*. 48:1–36.

Roth B, Becker N, Romeyke S, Schäfer S, Domnick F, Spinath FM. 2015. Intelligence and school grades: a meta-analysis. *Intelligence*. 53:118–137.

Salthouse TA. 1996. The processing-speed theory of adult age differences in cognition. *Psychol Rev*. 103:403–428.

Schermelleh-Engel K, Moosbrugger H, Müller H. 2003. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol Res*. 8:23–74.

Schubert A-L, Hagemann D, Frischkorn GT. 2017. Is general intelligence little more than the speed of higher-order processing? *J Exp Psychol Gen*. 146:1498–1512.

Schwaighofer M, Fischer F, Bühner M. 2015. Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educ Psychol*. 50:138–166.

Scott KM, Kline M. 2019. Enabling confirmatory secondary data analysis by logging data checkout. *Adv Methods Pract Psychol Sci*. 2:45–54.

Sheppard LD, Vernon PA. 2008. Intelligence and speed of information-processing: a review of 50 years of research. *Personal Individ Differ*. 44:535–551.

Simpson-Kent IL, Fuhrmann D, Bathelt J, The CALM Team, Kievit RA. 2019. Cross-sectional evidence for age differentiation-dedifferentiation in crystallized and fluid intelligence and their white matter substrates during childhood and adolescence. *BioRxiv*, 593509.

Sisk CL, Zehr JL. 2005. Pubertal hormones organize the adolescent brain and behavior. *Front Neuroendocrinol*. 26:163–174.

Tamnes CK, Roalf DR, Goddings A-L, Lebel C. 2018. Diffusion MRI of white matter microstructure development in childhood and adolescence: methods, challenges and progress. *Dev Cogn Neurosci*. 33:161–175.

Tucker-Drob EM. 2009. Differentiation of cognitive abilities across the lifespan. *Dev Psychol*. 45:1097–1118.

Vestergaard M, Madsen KS, Baaré WF, Skimminge A, Ejersbo LR, Ramsøy TZ, Gerlach C, Åkeson P, Paulson OB, Jernigan TL. 2011. White matter microstructure in superior longitudinal fasciculus associated with spatial working memory performance in children. *J Cogn Neurosci*. 23:2135–2146.

Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, Pérez-Stable EJ, Riley WT, Bloch MH, Conway K et al. 2018. The conception of the ABCD study: from substance use to a broad NIH collaboration. *Dev Cogn Neurosci*. 32:4–7.

Vollmar C, O'Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, Duncan JS, Richardson MP, Koepp MJ. 2010. Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *NeuroImage*. 51:1384–1394.

Wagenmakers E-J, Farrell S. 2004. AIC model selection using Akaike weights. *Psychon Bull Rev*. 11:192–196.

Wagenmakers E-J, Wetzels R, Borsboom D, van der Maas HLJ, Kievit RA. 2012. An agenda for purely confirmatory research. *Perspect Psychol Sci*. 7:632–638.

Wandell BA. 2016. Clarifying human white matter. *Annu Rev Neurosci*. 39:103–128.

Wechsler D. 2011. *Wechsler abbreviated scale of intelligence*. 2nd ed. San Antonio (TX): Pearson.

Weston SJ, Ritchie SJ, Rohrer JM, Przybylski AK. 2018. Recommendations for increasing the transparency of analysis of pre-existing datasets. PsyArXiv.