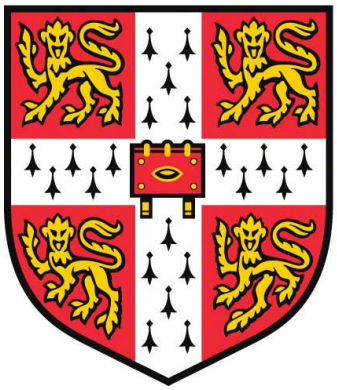
Theory-based Explorations of Associations between Human Brain Structure and Intelligence from Childhood to Early Adulthood

Ivan Levi Simpson-Kent

This dissertation is submitted for the degree of

Doctor of Philosophy



Wolfson College

MRC Cognition and Brain Sciences Unit

School of Clinical Medicine

University of Cambridge

March 2021

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as explicitly declared here and specified in the text. No content of this thesis plagiarises or is substantially the same as any work that has been previously submitted for a degree and/or publication(s) except where explicitly declared here and specified in the text. This dissertation does not exceed the prescribed 60,000-word limit for the Degree Committee of the Faculties of Clinical Medicine and Clinical Veterinary Medicine.

Chapter One's section **Jumping through the Developmental Windowpane** is partially adapted from:

Kievit R. A., **Simpson-Kent I. L.** (2021). It's About Time: Towards a Longitudinal Cognitive Neuroscience of Intelligence. *(due to appear in The Cambridge Handbook of Intelligence and Cognitive Neuroscience, Cambridge University Press, June 2021)*.

Chapter Two has been published (although appears in this thesis in a modified form to ensure consistent formatting and thesis coherence):

Simpson-Kent I. L., Fuhrmann D., Bathelt J., Achterberg J., Borgeest G. S., the CALM Team, Kievit R. A. (2020). Neurocognitive reorganization between crystallized intelligence, fluid intelligence and white matter microstructure in two age-heterogeneous developmental cohorts. *Developmental Cognitive Neuroscience, Special Issue: Flux 2018: Mechanisms of Learning & Plasticity*. doi: 10.1016/j.dcn.2019.100743

Chapter Three is available as a preprint (although appears in this thesis in a modified form to ensure consistent formatting and thesis coherence):

Simpson-Kent I. L., Fried E.I., Akarca D., Mareva S., Bullmore E.T., the CALM Team, Kievit R. A. (2021). Bridging brain and cognition: A multilayer network analysis of brain structural covariance and general intelligence in a developmental sample of struggling learners. Accepted, *Journal of Intelligence*. Preprint: *Biorxiv*. doi: 10.1101/2020.11.15.383869

Chapter Four is in progress for submission to a journal for publication

Acknowledgements

First and foremost, I want to acknowledge my three mothers, Barbara, Francine, and Mary. Without their sacrifices and guidance, I would never have been able to realise my potential and make it this far on my journey. I love you, Mommies.

Perhaps the most influential teacher I have had was Dr. John Berna, whose AP Psychology class in high school first got me interested in neuroscience. Your seemingly effortless lecturing made concepts as complex as the human mind appear tangible. Thank you for introducing me to my life passion.

I also want to thank my primary supervisor, Rogier Kievit, who has mentored me ever since I arrived in Cambridge during the autumn of 2017, lost in the city and in life. He has provided me with countless discussions that have sparked new ideas in me and has given me the freedom to pursue those ideas once they were properly formulated. Moreover, he has supported me throughout my bouts with mental illness. Without such support, I might not be here writing this sentence. I am also grateful for the Lifespan Cognitive Dynamics Lab and the MRC Cognition and Brain Sciences Unit where I have had the privilege of conducting my research. Even though you all never appreciate the true genius of my jokes, I appreciate the conversations and constructive critiques you have offered me throughout my tenure. I also want to thank my secondary supervisor, Joni Holmes, who has been patient with me during my rants about academia while writing up this dissertation. Speaking of academia, I would like to thank the University of Cambridge, Wolfson College, and the Cambridge Trust, whose facilities and resources created the opportunity for me to complete my doctoral studies.

I also want to acknowledge my friends, especially Siobhan, M, Andrea, Sinead, and Teuru, who were there for me during the most troubling time of my PhD. I thank Moataz, Silvana, and Sinead for proofreading parts of this thesis. I also want to thank Alisa, with the lessons I have learned from you, I am finally starting to discover some of the goodness and beauty in myself that you first saw when we met. Last but certainly not least, I want to thank my partner, Valerie, in such a short time you have made such a positive impact on my life.

This is a dissertation about human intelligence, but to complete it, I had to become well-versed in (life) resilience. The adversity I experienced in West Philadelphia and beyond has helped shape me into the person I am today. My triumphs have come with trauma, but triumphed I have.

Dedication

This dissertation is dedicated to all those past, present, and future whose mind is simultaneously their greatest asset, as well as their greatest adversary

Table of Contents

Abstract

Intelligence is often defined as the ability of an agent to learn, adapt to its environment, and solve novel challenges. However, despite over 100 years of theoretical development (e.g., general intelligence), widespread explanatory power (up to 50% of variance in cognitive scores), and the ability of intelligence measures to predict important life outcomes such as educational achievement and mortality, the exact configuration and neural correlates of cognitive ability remain poorly understood. This dissertation aims to make progress in this pursuit by exploring how human brain structure and intelligence correlate and co-develop with each other from childhood to early adulthood (ages 5 – 22 years). This endeavour is undertaken in three large cohorts (N range: 337 – 2072), guided by theory (e.g., crystallised and fluid intelligence), and implemented using rigorous, cutting-edge quantitative methods (i.e., structural equation modelling and network science). The results of this research provide robust evidence that the brain-behaviour relationships in intelligence are complex (i.e., consists of many independent yet interacting parts) and change nonlinearly during development.

The first study sought to elucidate the factorial structure and white matter substrates of child and adolescent intelligence using two cross-sectional, developmental samples (CALM: N = 551 (N = 165 neuroimaging), age range: 5 – 18 years; NKI-Rockland: N = 337 (N = 65 neuroimaging), age range: 6 – 18 years). In both samples, it was found (using structural equation modelling (SEM)) that cognitive ability is best modelled as two separable yet related constructs, crystallised and fluid intelligence, which became more distinct (i.e., less correlated) across development, in line with the age differentiation hypothesis. Further analyses revealed that white matter microstructure, most prominently of the superior longitudinal fasciculus, was strongly associated with crystallised (*gc*) and fluid (*gf*) abilities. Finally, SEM trees, which combines traditional SEM with decision trees, provided evidence for developmental reorganisation of *gc* and *gf* and their white matter substrates such that the relationships among these factors dropped between ages 7 – 8 years before increasing around age 10. Together, these results suggested that shortly before puberty marks a pivotal phase of change in the neurocognitive architecture of intelligence.

The second study builds upon the first by again examining the neurocognitive structure of intelligence, this time from a network perspective. The network or mutualism

theory of intelligence presupposes direct (statistical) interactions among cognitive abilities (e.g., maths, memory, and vocabulary) throughout development. Therefore, this project used network analytic methods (specifically graphical LASSO) to simultaneously model brain-behaviour relationships essential for general intelligence in a large (behavioural, N = 805; cortical volume, N = 246; fractional anisotropy, N = 165), developmental (ages 5 – 18 years) cohort of struggling learners (CALM). Results indicated that both the single-layer (cognitive or neural nodes) and multilayer (combined cognitive and neural variables) networks consisted of mostly positive, small partial correlations, providing further support for the mutualism/network theory of cognitive ability. Moreover, using community detection (i.e., the Walktrap algorithm) and calculating node centrality (absolute strength and bridge strength), convergent evidence suggested that subsets of both cognitive and neural nodes play an intermediary role 'between' brain and behaviour. Overall, these findings suggest specific behavioural and neural variables that may have greater influence among (or might be more influenced by) other nodes within general intelligence.

The final study investigated the longitudinal relationships between human cortical grey matter structure and measures of decision-making, risk-related behaviours, and spatial working memory from adolescence to early adulthood (ages 14 – 22 years). In the IMAGEN study (maximum N across time points/waves = 2072), latent growth curve models were used to estimate the baseline and longitudinal associations between behavioural measures and cortical surface area, thickness, and volume. Univariate models (only behavioural or neural measures) revealed that performance in decision-making, risk-related behaviours, and spatial working memory, as well as brain structure changed nonlinearly from mid-adolescence (age 14) to early adulthood (age 22). Furthermore, bivariate models (combined behavioural and neural measures) provided evidence for adaptive reorganisation (behaviour intercept predicts changes in brain structure) but not structural scaffolding (brain structure intercept predicts changes in behaviour). Furthermore, findings suggested that there were no correlated changes between behavioural and brain structure slopes (rates of change from mid-adolescence to early adulthood).

This dissertation concludes by summarising the core results, addressing key limitations, and discussing avenues for future research. Taken together, this thesis hopes to convince cognitive neuroscientists that to understand cognitive ability and its neural

determinants, they (we) must work more diligently toward building *coherent*, *rigorous*, and *testable* neurocognitive theories of intelligence—particularly under the conceptual and analytic paradigm of complex systems.

Chapter One

Laying the Foundations:

Neurocognitive Paradigms for Understanding

Human Intelligence

The section **Jumping through the Developmental Windowpane: Longitudinal Studies of Concurrent Structural Brain and Cognitive Changes from Childhood to Early Adulthood** from this chapter is partially adapted from:

Kievit R. A., **Simpson-Kent I. L.** (2021). It's About Time: Towards a Longitudinal Cognitive Neuroscience of Intelligence. *(due to appear in The Cambridge Handbook of Intelligence and Cognitive Neuroscience, Cambridge University Press, June 2021).*

My contributions to the above publication are:

1. Co-conceptualisation and planning (with R. A. Kievit) about content of publication.
2. Co-selection (with R. A. Kievit) of relevant literature to include in the publication.
3. Co-writing and editing (with R. A. Kievit), including confirmation of final version, of the chapter manuscript for publication in *Cambridge University Press*.

All other sections of Chapter One were led and conceptualised solely by me.

Spearman's Positive Manifold of Cognitive Abilities and the *g* factor of Individual Differences in General Intelligence

Intelligence is often defined as the ability of an agent to learn, adapt to its environment, and solve novel challenges (e.g., see Gottfredson, 1997 and Legg and Hutter, 2007). It has been over 100 years since Charles Spearman published his landmark observations that, across domains (e.g., language abilities, mathematics and music), individuals who scored higher on one cognitive task also tended to score higher on all other cognitive tasks, in relation to their peers (Spearman, 1904). He termed his discovery the 'positive manifold' of cognitive abilities (as all abilities positively correlated with each other, top of Figure 1). Spearman attempted to explain his results by proposing the general or ***g* factor** to account for the *shared covariance* among individuals on cognitive tasks. Later on, he also postulated the existence of specific or ***s* factors**, which described the *unique variance* among individuals (Spearman, 1927). Taken together, Spearman's "two-factor" (*g* and *s*) theory, commonly referred to collectively as the *g factor* model of human intelligence (Figure 1, bottom), states that differences in cognitive performance are jointly due to individuals possessing a lower or higher *g*, which relates to general cognitive abilities such as abstract reasoning and problem-solving, as well as possessing lower or higher levels of specific factors (*s*) that enable them to outperform their counterparts on more narrow cognitive tasks (e.g., mathematics, music, etc., see Spearman, 1927). Under Spearman's theory, individual differences in general intelligence are caused by the combination of general (*g*) and specific (*s*) factors, both of which vary from person to person. Moreover, *g* is assumed to account for most of the variance in general cognitive ability (e.g., innate ability through genetics) while *s* explains individual differences in narrow abilities (task-specific variance).

	Maths	Verbal Ability	Reasoning
Maths	—		
Verbal Ability	.78	—	
Reasoning	.69	.65	—
g	**.90**	**.77**	**.95**

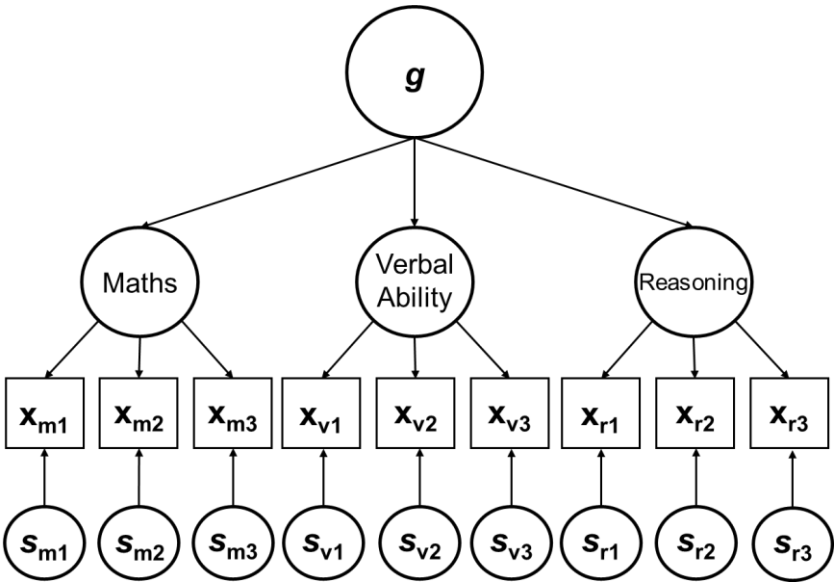


Figure 1. Top: Example correlation matrix showcasing the positive manifold of cognitive abilities. Bottom row shows factor loadings from *g* to cognitive tasks. Bottom: *g* factor model of human intelligence including general (*g*) and specific (*s*) factors. Note that error (residual) terms are also estimated in this model but are not shown. x's denote specific instances/measurements of cognitive abilities (e.g., x_{m1} = arithmetic)

Across intelligence datasets, Spearman's *g* factor model typically accounts for between 20 – 50% of the total variance in individual differences in cognitive ability (Deary et al., 2010), as well as reliably predicts important life outcomes such as educational and occupational achievement (Deary et al., 2007; Hegelund et al., 2018) and mortality (Calvin et al., 2011). Today, the *g* factor is recognised as one of the most replicable findings in psychology and is not limited to WEIRD (Western, Educated, Industrialised, Rich, and Democratic, see Henrich et al., 2010) samples, which are overrepresented in psychological studies, as *g* has been further verified in non-Western populations (see Warne and Burningham, 2019). Moreover, the *g* factor has been found in non-human animals such as dogs (Arden and Adams, 2016),

orangutans (Damerius et al., 2019), mice and primates (Locurto and Scanlon, 1998; Reader et al., 2011), and has even been extended to include performance of human groups (*c* factor of 'collective intelligence', Woolley et al., 2010; 2015)[1]. Together, these studies strongly suggest that a single cognitive ability factor (whether it be in humans, other organisms, or groups) has extraordinary predictive power when examining a range of abilities and outcomes.

Despite the efficacy of the *g* factor model to account for individual differences in cognitive ability, several other theories have been proposed, which give alternate ontologies of the positive manifold (rather than exclusively *g*). Most of these competing formulations posit the presence of group factors (e.g., crystallised and fluid intelligence, see Cattell, 1967, 1963) that cluster abilities measured by tests with similar properties (e.g., verbal ability and working memory, see McGrew, 2009; Schneider and McGrew, 2012). Other models conceptualise the positive manifold (and hence *g*) as arising from network interactions among cognitive abilities during a given task or over the course of early development (e.g., childhood, see van der Maas et al., 2006) while others view *g* as the (population-level) emergent property of within-individual differences in executive processes such as working memory capacity (Kovacs and Conway, 2016). In the next section, I will briefly describe these theory types—specifically ***gc-gf* theory**, **Cattell-Horn-Carroll (CHC) theory**, **mutualism**, and **Process Overlap Theory**—and evaluate how they relate to Spearman's *g* theory.

[1] Interestingly, *c* has been found to be only moderately correlated with average or maximum individual intelligence. Instead measures related to "social perceptiveness" (e.g., Reading the Mind in the Eyes (RME) Test, Baron-Cohen et al., 2001) and "moderate level of cognitive style diversity" (Aggarwal et al., 2015) are better predictors of *c*.

Going Against the (*g*)rain:

Refinements to the *g* factor Theory of General Intelligence

Crystallised and fluid intelligence:

Cattell's *gc-gf* theory

As described above, the *g* factor is highly effective in accounting for individual differences on cognitive scores (accounts for up to 50% of the total variance) and predicting important life outcomes. As a result, it provides a parsimonious framework (as it only assumes one general factor) for understanding how and why humans as well as other organisms perform well on tests of cognitive ability, relative to their peers. However, since Spearman first hypothesised *g*, several researchers, having conceptual as well as theoretical issues with his theory, have proposed alternative theories that have become prominent in the intelligence literature.

One of these refinements to Spearman's *g* theory came from his student, Raymond Cattell. According to Cattell, rather than of a single factor underlying all cognitive ability, general intelligence was instead composed of two distinct yet correlated constructs or types, which he labelled crystallised (***gc***) and fluid (***gf***) intelligence (Cattell, 1963; 1967). Cattell suggested that *gc* represents the capacity to effectively complete tasks relying on experience and knowledge obtained (mostly) from schooling (e.g., arithmetic, vocabulary, etc.), whereas he defined *gf* as an individual's capacity to solve novel problems devoid of task-specific knowledge, instead using abstract reasoning and pattern recognition (also see Deary et al., 2010). Hence, Cattell replaced Spearman's *g* with his *gc-gf* theory (Figure 2) as the core psychological mechanisms that give rise to the positive manifold. In Chapter Two, I statistically compare Spearman's *g* factor to Cattell's *gc-gf* theory and investigate their structural neural correlates (via white matter fractional anisotropy) in two large developmental samples of children and adolescents (Centre for Attention, Learning and Memory, N = 551; NKI-Rockland, N = 337).

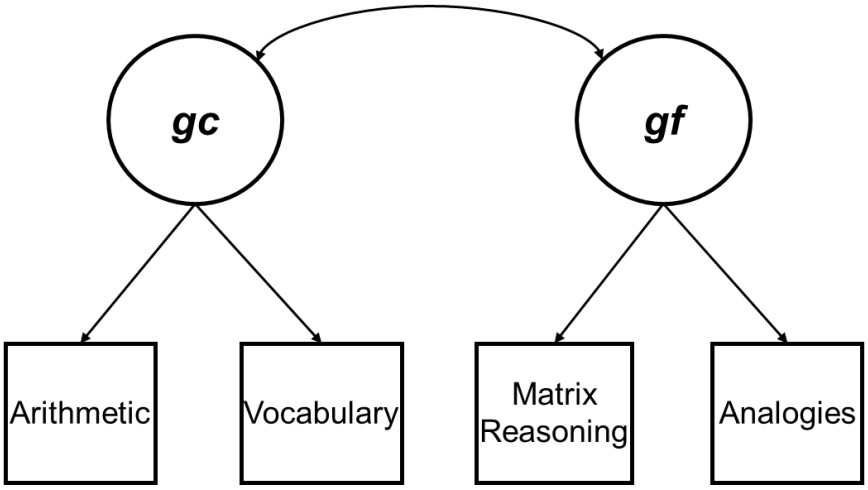


Figure 2. Cattell's *gc-gf* factor model of human intelligence. Here, *gc* and *gf* are conceived to cause changes in the factor loadings of crystallised (Arithmetic and Vocabulary) and fluid (Matrix Reasoning and Analogies) tasks, respectively. Note, in this model, *gc* and *gf* are correlated with each other (double-headed arrow).

Let's get it stratum:

Cattell-Horn-Carroll (CHC) theory

In addition to Cattell's *gc-gf* theory of intelligence, other 20th century theorists put forth factor models hoping to elucidate the nature of the positive manifold. These include (but are not limited to) the Cattell-Horn Extended *gc-gf* theory (Horn and Cattell, 1966), the bifactor model (Holzinger and Swineford, 1937; also see Watkins and Beaujean, 2014 for illustration) and Carroll's three-stratum theory (Carroll, 1993; 1997). Arguably the most comprehensive model of intelligence to date (see Flanagan and Dixon, 2014) is the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (McGrew, 2009; Schneider and McGrew, 2012), which combines the formulations of Cattell, Horn, and Carroll into a single hierarchical model.

According to CHC theory (Figure 3), general intelligence (*g*) sits at the apex of the hierarchy of cognitive abilities and is considered to be a **'General Ability or Factor' (stratum III)**, analogous to Spearman's original conception of *g*. Beneath *g* comes the **'Broad Abilities' (stratum II)** that include specific *groups* of cognitive abilities such as crystallised (*gc*) and fluid intelligence (*gf*) mentioned above as well as (but not limited to) short-term memory (***gsm***), mathematical knowledge needed for manipulating numerical symbols (***gq***), and processing speed (**gs**), which is required to quickly and accurately manipulate information such as numbers and to read sentences. As with *gc-gf* theory, the broad abilities represent 'group factors' that organise related tasks (e.g., verbal abilities such as reading and spelling) so they

are separate from more dissimilar tasks such as arithmetic and mathematical problem-solving (Maths). Although Spearman included *s* factors in his two-factor model to account for fine-gained individual differences in specific abilities, he was aware that group factors might be necessary if the cognitive content of different tests significantly overlapped (e.g., see Thomson, 1947). Lastly, at the lowest level rest the **'Narrow Abilities' (stratum I)**, which refer to even more specific abilities than those of stratum II. In describing stratum I, Carroll stated that these abilities ''represent greater specializations of abilities, often in quite specific ways that reflect the effects of experience and learning, or the adoption of particular strategies of performance'' (Carroll, 1993, p. 634). This depiction can be compared with Spearman's *s* factors in the sense that their contributions depend on the task and training level of the individual.

The CHC theory has garnered widespread empirical and statistical support (Benson, 2008; Flanagan, 2000; see Flanagan and Dixon, 2014 and McGrew, 2009 for overviews but also see Canivez and Youngstrom, 2019 for conflicts regarding CHC theory and its applications) and is a deeply influential paradigm for thinking about cognitive abilities in the cognitive psychology community. In closing, while the aforementioned models and their subsequent iterations differ from each other in terms of the constructs involved (e.g., *g* only, *gc* and *gf* only, or combined with additional factors) and statistical properties (i.e., correlational, orthogonal, and/or hierarchical structure), a comprehensive comparison of these models is beyond the scope of this dissertation (but see Beaujean, 2015 and Flanagan and Dixon, 2014 for reviews).

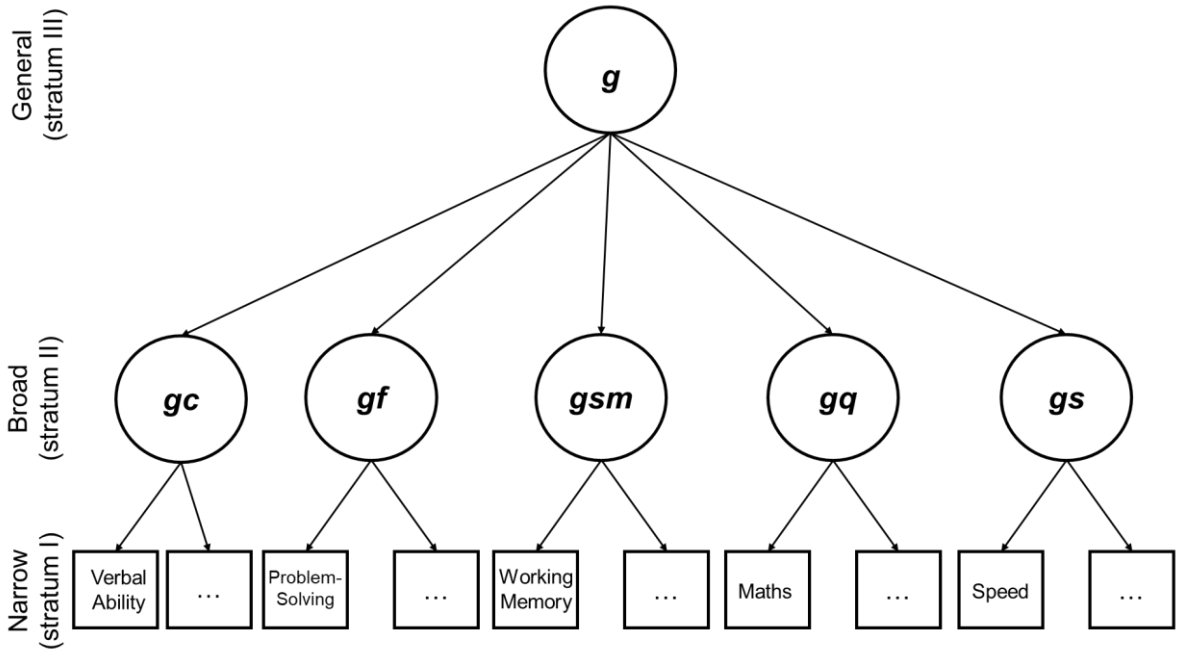


Figure 3. The Cattell-Horn-Carroll (CHC) theory of cognitive abilities. General intelligence (*g*) is located at the top (stratum III), followed by Broad abilities (stratum II), and bottoms out with Narrow abilities (stratum I). Abbreviation key: crystallised (*gc*) and fluid (*gf*) intelligence (both studied in Chapter Two), short-term memory (*gsm*), quantitative knowledge (*gq*) and processing speed (*gs*).

The emergence of *g* through interactions of cognitive abilities throughout development: The mutualism theory of general intelligence

An alternative perspective to traditional factor-model theories is to conceptualise general intelligence as the consequence of a dynamic network that evolves over time. This theory, known as mutualism, claims that the positive manifold results from *positive, reciprocal interactions* between cognitive abilities (van der Maas et al., 2006). Hence, early in cognitive development, cognitive abilities (e.g., vocabulary and reasoning) are weakly correlated with each other, resulting in little to no *g* factor/positive manifold. Instead, over time—for instance from early childhood until adolescence—these associations increase in strength and become more positive, eventually giving rise to the positive manifold and *g*. Mutualism (see Figure 4) has gained support from several studies (Ferrer and McArdle, 2004; Kan et al., 2019; Kievit et al., 2019, 2017) and reviews (Peng et al., 2019; Peng and Kievit, 2020) that have demonstrated that cognitive abilities such as vocabulary drive positive changes in other abilities (e.g., reasoning) and vice versa. Since its initial conception, the mutualism/network model of intelligence has been expanded and further formalised (van der Maas et al., 2021, 2017) to

incorporate additional psychometric theories (i.e., Spearman's *g* factor, Cattell's crystallised (*gc*) and fluid (*gf*) abilities, and the Dickens-Flynn gene-environment interaction model of IQ, Dickens and Flynn, 2001). Moreover, it has recently been further formalised to help explain cognitive development (Savi et al., 2019) and has even been applied to help understand sensitive periods in cognitive development (Kievit, 2020). In Chapter Three, I apply the mutualism/network model paradigm to include metrics of brain structure (i.e., grey and white matter).

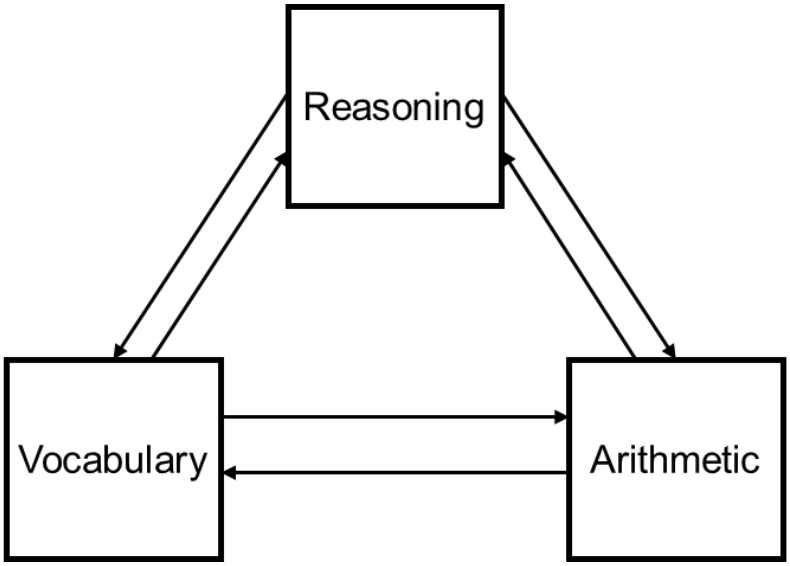


Figure 4. Simplified model of the mutualism theory of cognitive ability. Single-headed arrows indicate interactions (positive reciprocal correlations) between cognitive abilities (arithmetic, vocabulary, and reasoning).

Process Overlap Theory:

General intelligence as a formative, emergent phenomenon

One of the most recent psychometric theories of general intelligence (also see Savi et al., 2019) that attempts to elucidate the positive manifold is Process Overlap Theory (POT, Kovacs and Conway, 2016). POT attempts to elucidate interindividual (between-person) differences in general intelligence by including constraints on *g* in the form of intraindividual (within-person) differences. According to the authors, POT separates itself from previous accounts in "that it integrates psychometrics, cognitive psychology, and neuroscience" (p. 152). Thus, unlike the formal theories described above, POT incorporates properties of brain structure and function, especially insights obtained from the Parieto-Frontal Integration

Theory (Jung and Haier, 2007, see the section **The early years: The Parieto-Frontal Integration Theory (P-FIT) of intelligence** below) as well as studies of the neural correlates of *executive processes* (especially fluid intelligence and working memory), which together emphasise the *network* properties of the brain and their links with intelligence as opposed to only individual regions. I discuss the neural interpretation of POT further in the section **Overlapping processes in the brain: The neural determinants of Process Overlap Theory** below.

To build their theory, Kovacs and Conway, 2016 first lay out the three fundamental assumptions or "axioms" of their theory (p. 152 – 153). First and foremost, the *g* factor is a necessary statistical consequence of the positive manifold. In other words, in any correlation matrix with *solely positive values*, it is inevitable that a single (latent/unobserved) factor can be extracted using factor analytic techniques (see Krijnen, 2004 for technical analysis). Furthermore, this general factor will have positive correlations (indexed by factor loadings) between it and its manifest/observed indicators or first- or second-order factors if the model is hierarchical (e.g., CHC theory). Secondly, independent from this inescapable mathematical result, any such latent variable *must* be assumed to *causally influence* the positive correlations among the measured variables (known as entity realism, see Borsboom et al., 2003). Lastly, as "*differential* constructs" (p. 153), there are no one-to-one mappings between latent variable between-subject variation and within-person processes (Molenaar and Campbell, 2009; Voelkle et al., 2014 but also see Schmiedek et al., 2020 for recent example in intelligence data). A corollary of this is that latent variables such as *g* do not exist in isolation, but instead depend on individual differences (variation in cognitive ability among multiple people). Therefore, *g* is necessarily a between-subject (population-level) phenomenon. To quote the authors' explanation:

> "…the last survivor of a meteor collision with Earth would still have cognitive abilities and mental limitations but would not have *g*…Hence the scope of any explanation of the positive manifold, including but not restricted to latent variables, is not necessarily directly applicable to single individuals." (p. 153)

With these axioms, Kovacs and Conway put forth Process Overlap Theory (POT), which ties together robust evidence that implicates the overlapping roles of domain-general—for instance, executive processes such as working memory capacity (WMC, see Burgoyne et al., 2019; Conway et al., 2003, 2002; Engle et al., 1999 for studies exploring the relations between

WMC and general (fluid) intelligence)—and domain-specific abilities that are recruited when individuals perform narrow tasks (e.g., verbal reasoning). The authors surmise that "Such a pattern of overlap of executive and specific processes explains the positive manifold as well as the hierarchical structure of cognitive abilities" (p.161, Kovacs and Conway, 2016). In other words, the reason many complex abilities such as WMC strongly correlate with measures such as fluid intelligence (*gf*) is that both involve a "multicomponent construct with overlapping processes" (p. 162). This also explains why more specific, elementary tests such as simple span tasks do not significantly correlate with each other and explain unique sources of variation in *g* (i.e., they do not share overlapping processes).

Under POT (Figure 5), *g* is the emergent (formative latent variable) result of a finite number of distinct abilities rather than a single, unitary entity responsible for the shared variation among cognitive tasks (reflective latent variable). However, below *g*, specific abilities take on a reflective latent variable structure and are correlated with each other to formalise the overlapping process (i.e., not completely independent) that underlie them. In addition to arguing that *g* is an emergent phenomenon, POT claims that executive processes (e.g., WMC) function as limiting factors for successful completion of tasks requiring domain-general abilities. For example, if a participant lacks sufficient working memory or inhibitory control, they will not be able to perform well on goal-related tasks (tapping *g*) such as playing a video game that consists of remembering rules (e.g., sequence of steps needed to pass a level) and avoiding obstacles (e.g., game-generated monsters that try to kill you before you finish a level). Moreover, these participants would also struggle on domain-specific tasks (e.g., gaining enough points to unlock a single item necessary to continue through a level).

Kovacs and Conway mathematically formalised their theory (Equation 1 of Kovacs and Conway, 2016) in an item response model that estimates the "probability of a person (p) arriving at a correct answer on a test item (i) that taps component processes (C) from a number of different domains (D)" (p. 162). The model assumes that each cognitive dimension is relatively (ontologically) independent from each other. POT is formalised as a hybrid multidimensional item response model (see Reckase, 2009) with both compensatory (additive sums within cognitive domains) and noncompensatory (products of across-domain abilities) elements. Given that across-domain abilities such as executive processes act as bottlenecks to higher-order *g*, this makes p "a nonadditive and nonlinear [product] function of the score on

each individual dimension" (p. 163). Therefore, a person with below average capacity in one executive process such as working memory capacity will have a low probability of successfully performing a specific working memory tasks, even if they display high scores for that specific test. This is because their deficiency in *nonlinear* WMC (e.g., complex span) cannot be offset by *additive* satisfactory or above average capacity in the easier working memory tests (e.g., simple span). In closing, POT formalises *g* as a *population-level* phenomenon that *emerges* (formative) from the *within-person* overlapping of (mainly executive) cognitive processes that are tapped by various psychometric tests. Now, having described several prominent psychological models of intelligence, I next discuss how recent perspectives and findings from cognitive neuroscience have further informed understanding of intelligence.

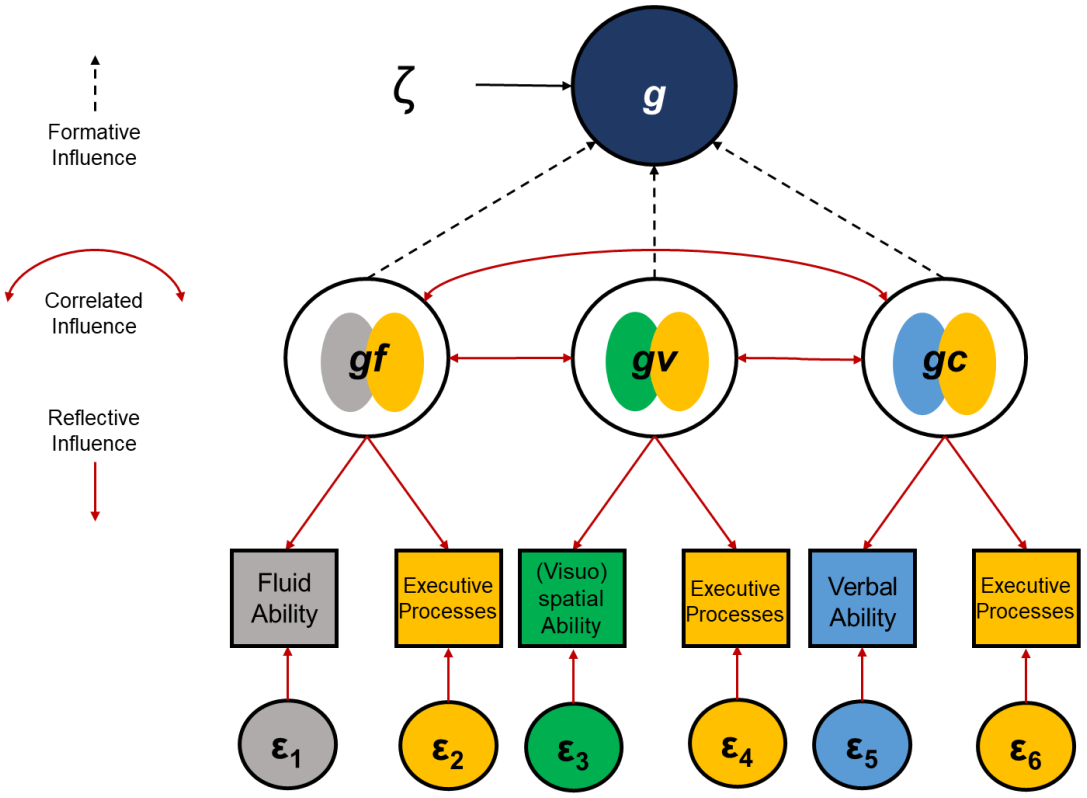


Figure 5. Simplified depiction of Process Overlap Theory (POT) as a latent variable model. Single-headed arrows (solid red: reflective, black dashed: formative) indicate causal directions among types of cognitive tests (boxes), latent constructs (circles), and error/residual terms (ε). Double-headed arrows indicate correlations between cognitive domains. General intelligence/the positive manifold (*g*) is an emergent property of interactions of theorised cognitive abilities (here *gf*, *gv* and *gc*) from lower stratum. The ζ term represents residual/unexplained variance (e.g., neural and environmental processes) that also contribute to the emergence of *g*. Note that 'Fluid Ability' is also an executive process but is coloured grey to distinguish it from other executive process (e.g., WMC) more widely involved in the other latent ability factors (i.e., *gv* and *gc*).

From Psychometrics to Cognitive Neuroscience:

The Search for the Neural Foundations of General Intelligence

In addition to the psychological literature, the field of cognitive neuroscience has provided an emerging source of insight in uncovering the neural causes of individual differences in general intelligence. These cognitive neuroscience approaches go beyond purely behavioural descriptions and formulations by also measuring the neural process necessary for cognitive ability. In this section, I will describe five leading neurocognitive theories of general intelligence (in chronological order in which they appear in the literature): **Parieto-Frontal Integration Theory**, **the multiple-demand system**, **Process Overlap Theory**, **The watershed model** of individual differences in fluid intelligence, and **The Network Neuroscience Theory of Human Intelligence**.

The early years:

The Parieto-Frontal Integration Theory (P-FIT) of intelligence

The first comprehensive attempt to propose a neural basis of intelligence was Jung and Haier's Parieto-Frontal Integration Theory (P-FIT, Jung and Haier, 2007). P-FIT, formulated based upon a review and synthesis of many neuroimaging studies on cognitive ability, asserts that individual differences in intelligence derive from information exchange primarily between inferior and superior parietal (Brodmann areas (BAs): 7, 39, and 40) and dorsolateral prefrontal (BAs: 6, 9, 10, 45, 46, and 47) cortical areas, but also regions of the occipital (BAs: 18 and 19) and temporal (BAs: 21 and 37) lobes, as well as the anterior cingulate (particularly BA 32). Moreover, Jung and Haier proclaimed that white matter tracts, especially the arcuate fasciculus but also the superior longitudinal fasciculus, assist in producing intelligent behaviour by integrating informational content from individual grey matter regions through anatomical linkages. They assumed that occipital (e.g., the fusiform gyrus (BA 37) for the analysis of visual stimuli) and temporal (i.e., Wernicke's area (BA 22) for auditory content) regions played an especially pronounced role in early processing of relevant information (i.e., a sensory stimulus). The results of this initial stage are then "fed forward" (p. 138) to parietal areas such as the angular (BA 39), superior parietal (BA 7) and supramarginal (BA 40) gyri for additional processing. Lastly, parietal regions interact with frontal cortices to "test various

solutions to a given problem" (p. 138), eventually leading to a final decision made in the anterior cingulate.

The acclaim of P-FIT theory is well-founded and justified. It represents a tour de force of scientific investigation, the culmination and synthesis of 37 studies (total N = 1,557 participants) using various neuroimaging techniques (e.g., functional MRI, positron emission tomography, diffusion tensor imaging, etc.) and measures of intelligence (chess performance, analogy and maths reasoning, WAIS vocabulary, etc.), and even incorporating findings from lesion and genomics studies. Despite this, however, Jung and Haier never intended for the P-FIT to solve the quest to understand intelligence. Instead, they argued that "the model provides a framework for testing new hypotheses in future experimental designs" (p. 135). Therefore, progress was still needed and welcomed despite their impressive achievement.

Neural *g*?

The multiple-demand system underlying complex problem-solving

Shortly after Jung and Haier's publication of P-FIT Theory, Duncan, 2010 presented his *g*-style theory of human intelligence, known as the brain's multiple-demand (MD) system. The MD system (also see Fedorenko et al., 2013), also sometimes referred to as the frontoparietal control network (Vincent et al., 2008; see Uddin et al., 2019 for additional functional network names association with MD ("Lateral Frontoparietal Network") regions), is thought to be involved in a wide array of cognitive processes, including but not limited to fluid intelligence (*gf*), a form of abstract reasoning discussed above. Furthermore, the MD system is only activated (the MD system has been studied mostly using fMRI, but in principle could be assessed using other functional modalities such as EEG, PET or MEG) whenever an individual performs a challenging task, and is, therefore, usually studied using an easy/hard task contrast. Composed primarily of frontal and parietal brain regions, the MD system acts as a functional network that, through its coordinated activity, enables intelligent behaviour. According to Duncan, 2010, the MD system's principal function pertains to the solving of complex tasks by first partitioning them into smaller, simpler sub-tasks. This "sequential mental programming" (p. 177) allows for easier solvability and facilitates learning. To summarise, supporters of the MD system argue that domain-general cognition can (in large

part) be explained by a *unitary core* (i.e., mainly frontal and parietal regions) of functional brain activity (see Figure 6).

While early work on the MD system focused on fluid intelligence and complex problem-solving (see Duncan et al., 2020 for recent review of how the MD system integrates 'distributed' brain activity to produce 'organised' cognition), it is known that *gf* represents only one facet of overall cognitive ability. In other words, the MD system, while responsible for higher-level complex problem-solving on challenging tasks, might play a smaller role in more elementary cognitive areas such as language abilities. One such study has recently supported this hypothesis (Woolgar et al., 2018). Using probabilistic volume maps obtained from MRI scans in 80 patients with either frontal or posterior lobe lesions, they found that MD-weighted, but not language-weighted, lesion volume (negatively) predicted change in fluid intelligence (lower scores represented greater deficit after injury). Moreover, a distinct effect was seen in the prediction of verbal fluency scores such that language-weighted rather than MD-weighted lesion volume was a significant (negative) indicator of change in verbal fluency (lower scores indicated greater postmorbid language deficit)[2]. The authors concluded that their results supported the notion that language (and the brain systems that support it) are not essential for complex thought. However, it is the case that many complex tasks, including but not limited to the translation of challenging academic texts and memory sports competitions, also tap into fluid-like cognitive resources to form associations (e.g., between concepts in different languages or dialects and mnemonic devices) that involve verbal (language) abilities. Therefore, an alternative interpretation could be that the MD and language systems are instead part of a larger conglomerate of brain regions that, although specialised with respect to their intrinsic function, work together to produce cognitive ability in all its diverse forms.

[2] Along these lines, a more recent study (Diachek et al., 2020) found similar results in a larger sample (N = 481) collected from datasets of 30 "word and sentence comprehension experiments" (p. 4537). Specifically, the left MD-system ROIs responded stronger to verbal tasks, indicating language lateralisation. In addition, the MD-system displayed weaker activation during sentence and 'passive' comprehension tasks, in contrast to the 'language-selective network'. Therefore, the MD itself seems to be further specialised into functional units.

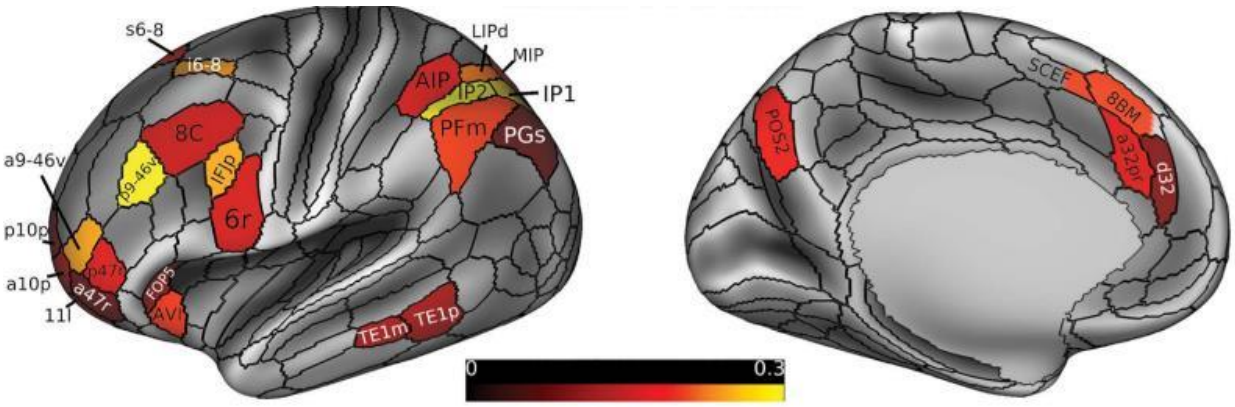


Figure 6. The 'Extended' multiple-demand (MD) system represented by non-grey brain regions. This figured is based on functional MRI and reproduced from Assem et al., 2020 with permission.

Overlapping processes in the brain:

The neural determinants of Process Overlap Theory

Under Process Overlap Theory (POT, Kovacs and Conway, 2016), the primary focus is on executive processes such as fluid intelligence (*gf*) and working memory capacity (WMC), which act as bottlenecks to forming general intelligence (*g*, see Figure 5). According to the authors, a large body of evidence (starting with P-FIT theory, Jung and Haier, 2007, see above) implicates the frontal lobe (and some parietal regions) as the section of cortex essential for *gf* and WMC. As part of the MD system/frontoparietal control network, the frontal lobe and especially the prefrontal cortex (PFC) functions as integration and regulatory hub that coordinates information exchange between it and more specialised regions (e.g., those responsible for specific abilities). Hence, the within-subject frontoparietal control network (mainly via frontal regions such as the PFC) represents the neural capacity limits/bottleneck in information processing (Dux et al., 2006; Koechlin and Hyafil, 2007; Marois and Ivanoff, 2005; Tombu et al., 2011) in a wide variety of cognitive demands and correspond to the overlapping executive processes that lead to between-subject differences in *g*.

Further support for this claim can be found in lesion and statistical modelling studies that link executive processes (e.g., *gf* and WMC) to imaging metrics from the frontal lobe. For instance, in 80 patients Woolgar et al., 2010 used multiple regression to estimate the extent to which damage to specific brain regions-of-interests (ROIs) correlated with deficit in *gf*. They found that, on average, damage localised to prefrontal and parietal areas of the MD system resulted in greater loss (i.e., 6.5 points) in fluid ability than if injury was elsewhere in the brain

(1 point). Moreover, using partial correlations (see Chapter Three for use of this statistical technique), the researchers found that, regardless of location within the brain, each ROI studied explained unique variance in fluid IQ loss.

Various multivariate techniques have been used to examine these questions. One such tool is structural equation modelling (see Chapter Two), which has been used to pinpoint the central influence of frontal areas in executive processes. For example, in 104 participants Román et al., 2014 examined the relationship between both general (e.g., *g*) and specific (e.g., crystallised (*gc*) and fluid ability (*gf*)) measures of intelligence and brain morphology, assessed by cortical surface area, thickness and volume. They found a 'reversed hierarchy' in the grey matter correlates of cognitive ability such that at higher-order, more general psychometric scales (e.g., general factors such as *g*) "the smaller the number of relevant gray matter clusters accounting for individual differences in intelligent performance" (p. 3805). In other words, as one moves from specific to general factors of intelligence, less total (but still predominantly frontal) brain structure is needed to explain variance in cognitive performance scores.

Lastly, Kievit et al., 2016 investigated the relation of fluid reasoning to brain activity using task-based fMRI, which enabled the authors to *simultaneously* determine and distinguish between individual differences (between-subjects) in fluid ability as well as performance dependent upon the level of difficulty (within-subjects) of the *gf*-related task. Using conjunction analysis (see Nichols et al., 2005) in a small sample (N = 34), Kievit and colleagues found three cortical areas that showed greater activation depending *both* on between-subject ability and level of difficulty: right middle and superior frontal gyri, bilateral angular gyri (superior parietal cortex), and bilateral precunei. They termed this convergence "local neural ergodicity" (p. 13), which has proven challenging to find in neuroscience data (e.g., Medaglia et al., 2011; Roberts et al., 2016; but also see Kievit et al., 2013 for a discussion of Simpson's paradox in psychology). Altogether, the agreement in (statistical) neural evidence among this and the above studies, despite using diverse methods, led Kovacs and Conway, 2016 to conclude that "The present state of research in neuroscience demonstrates that…the overlap the theory proposes appears to actually take place in the human brain" (p. 169).

The watershed model of individual differences in fluid intelligence

The watershed model of individual differences in fluid intelligence (Kievit et al., 2016; also see Penke et al., 2007) provides another example (as POT) of hierarchical representation and estimation of neurocognition. In a large (N = 555) adult lifespan (ages 18 – 87 years) sample (Cam-CAN, Shafto et al., 2014), Kievit and colleagues used a type of structural equation modelling formalisations known as MIMIC models (see Chapter Two for brief description and use of this model class) to explore the relations between processing speed, fluid intelligence, and white matter integrity. They found that a 'watershed' model, inspired by Cannon and Keller, 2006's attempt to conceptualise and quantify the complex and multilevel causes of mental disorders, showed better fit to the data than competing models (i.e., a single factor/unidimensional model).

The watershed model assumes 'hierarchical dependence', predicting that the most 'downstream' phenotype (in this case fluid ability) is the final result of a cascade of multiple 'upstream' causes (e.g., white matter integrity). Moreover, these factors only exert *indirect* influence on fluid intelligence via intermediate endophenotypes (Fornito and Bullmore, 2012), in this study tests of processing speed. This means that no causal pathway is drawn directly between, for example, specific white matter tracts and general (or fluid) cognitive ability. The watershed model also presupposes a *many-to-one* mapping of brain-behaviour relationships, known as degeneracy in neuroscience (Friston and Price, 2003 but also see Edelman and Gally, 2001 for how the concept more broadly applies to biological systems), as well as *increased statistical dimensionality* for more upstream compared to downstream levels due to them being more partially independent from each other. The performance of this model has been replicated (Fuhrmann et al., 2020) in two childhood and adolescent (ages 5 – 17 years) cohorts (CALM, Holmes et al., 2019, and NKI-Rockland, Nooner et al., 2012), again using processing speed (i.e., N-back task, rapid naming and trail-making) but also working memory tasks (e.g., forward and backward digit recall). In doing so, they extended the explanatory power (in terms of variance) of the model to include developmental periods prior to adulthood. Lastly, although so far only examined for fluid intelligence, uncontrolled eating (Garcia-Garcia et al., 2020) and nationalism (Zmigrod et al., 2018), the hierarchical watershed model can, in theory, also be applied to other cognitive domains such as language.

The dawn of modern network neuroscience:

The Network Neuroscience Theory of Human Intelligence

More recently, network neuroscience (Bassett and Sporns, 2017; Betzel, 2020) has evolved into a more quantitatively-oriented (e.g., increasingly using methods from computer science, mathematics, physics, etc.) and established subfield. Under this theoretical framework, the brain is conceptualised as a *complex system* of interconnected (i.e., cortical structure such as grey and white matter) and interactive (e.g., functional brain regions with significantly correlated co-activation patterns) networks that enable various cognitive abilities including sensorimotor autonomy and learning (Bassett et al., 2015, 2011; Bassett and Mattar, 2017). This view contrasts with previously dominant reductionist perspectives (see Kievit et al., 2011 for an overview of reductionism in cognitive neuroscience) that mostly focused on the *isolated* roles of individual brain regions (for example, primarily the prefrontal cortex) in contributing to cognition and other behaviours.

Network neuroscience employs techniques from network science (Barabási, 2016) such as graph theory and centrality analysis to characterise both universal and specific aspects of brain structure and function (Fornito et al., 2016). These methods can be applied across spatial (from single-cells to whole-brain), temporal (from milliseconds to across the lifespan), and topological (from individual nodes to global network) scales (Betzel and Bassett, 2017), as well as between species (van den Heuvel et al., 2016). Toward this end, several pioneering studies have revealed pervasive properties of brain network structure and function. These include, for example, the discovery of hubs (Sporns et al., 2007; van den Heuvel and Sporns, 2013), or nodes (e.g., individual brain regions) that share many connections with other nodes within the brain. Subsequent research has built upon this finding by identifying specific classes of hub regions such as the 'rich club' (Heuvel and Sporns, 2011), the 'diverse club' (Bertolero et al., 2017), and the 'flexible club', (Yin et al., 2020). Due to their numerous links (whether physically as in white matter or statistically through co-activation patterns found in fMRI), hubs help relay informational content throughout the brain.

Furthermore, nervous systems have been shown to exhibit modularity (Meunier et al., 2010; Sporns and Betzel, 2016), small-world topology (Bassett and Bullmore, 2006, 2017), and an economical trade-off between minimising neural wiring cost and maximising efficiency (Bullmore and Sporns, 2012). Together, these properties enable the brain to perform

both specialised (intra-modular) and global (inter-modular) processing and allow efficient information transfer to facilitate adaptive behaviours (e.g., intelligence). Lastly, concerted efforts have been made to formulate mathematical models to generate and mechanistically describe a variety of brain phenomena including (but not limited to) functional brain network organisation (Vertes et al., 2012), normal and abnormal neurodevelopment (Akarca et al., 2020; Vértes and Bullmore, 2015), and how connector hubs and modularity relate to cognition (Bertolero et al., 2018).

Drawing on these and other findings (further explored in Chapter Three), the latest network neuroscience proposal aiming to explain general intelligence (Spearman, 1904) is the Network Neuroscience Theory of Human Intelligence (Barbey, 2018). Barbey draws inspiration from four theories of intelligence: 1) Spearman's *g*, 2) Cattell's *gc-gf* theory and 3) the mutualism theory of cognitive ability (all explained above), and 4) Godfrey Thomson's sampling theory (Thomson, 1939, 1919, 1916). Thomson, one of Spearman's contemporaries, formulated his sampling theory of cognitive abilities and challenged Spearman's *g* as the causal mechanism underlying the positive manifold. According to Thomson, the myriad of positive correlations between tasks stem from a given test (e.g., vocabulary) *sampling* a large number of mental 'bonds' that each partially contribute to cognition. Therefore, if one were to devise assessments that *only* measure specific abilities—for example, the fundamental building blocks of simple vocabulary—the positive manifold would disappear. However, in practice most tests measure mental constructs that are too broad (e.g., fluid or verbal ability) and, therefore, share cognitive processes. Although there are arguments to the contrary (i.e., see Savi et al., 2019), Thomson's sampling theory, which states the positive manifold results from separate cognitive faculties each contributing to varying degrees to general cognitive ability, can be viewed as one of the first (and most forgotten) network perspectives and models of general intelligence.

Barbey, 2018 conceptualises *g* as a global network phenomenon that arises from the small-world typology, modularity, and dynamics of the brain. He finds correspondence between Thomson's sampling theory, Spearman's *s* factors and Cattell's *gc-gf* theory, and relates them to intrinsic connectivity networks (ICNs, Laird et al., 2011; Seeley et al., 2007). For instance, it is well-known that many neural regions are specialised for particular functions (e.g., Power et al., 2011), both for simple processes and cognition, which allows for increased

local efficiency and minimises wiring cost (Bullmore and Sporns, 2012). The functional and structural specialisation of ICNs parallels Spearman's *s* factors (specific abilities developed through practice) and Thomson's narrow sampling of cognitive domains (a test that only measures verbal ability). Due to their intra-modular connectivity and hub architecture (De Domenico et al., 2016; Hilger et al., 2017; Power et al., 2013; van den Heuvel and Sporns, 2013), the presence of ICNs (e.g., the default mode network) allows individuals to access prior knowledge stores (i.e., those obtained via education), typically a hallmark of crystallised intelligence. According to Barbey, fluid ability, the second component of Cattell's *gc-gf* theory (Cattell, 1963), arises from ICNs (e.g., cingulo-opercular and frontoparietal networks) that contain nodes with weaker inter-module ties, which is needed for collective behaviour (Schneidman et al., 2006). This enhances flexibility and the efficiency of global processing by forming new links between distinct modules (see Gallos et al., 2012), which in turn enable individuals to adapt to novel environments and situations (Barbey et al., 2014, 2012; Gläscher et al., 2010).

These characteristics of brain functional and structural organisation (also see Bertolero et al., 2015), which balance short and long-range connections through dynamic small-world typology, facilitate transitions (Betzel et al., 2016) between two primary network states (Gu et al., 2015): 1) 'easy-to-reach' (those used to access prior knowledge needed to carry out crystallised tasks), and 2) 'difficult-to-reach' (necessary for integrating information, giving rise to fluid-like cognitive abilities). Therefore, general intelligence (or *g*) is best understood as the emergent result (similar, in principle, to mutualism and Process Overlap Theory) of this 'tug of war' between specific and broad ICNs. Broad ICNs (especially frontoparietal networks) help to drive the brain to difficult-to-reach or random states due to the greater variability of their connections between modules compared to within them (Braun et al., 2015; Cole et al., 2013). Altogether, this constrained flexibility allows the brain to adapt to novel cognitive domains (e.g., in abstract reasoning), while preserving the retrieval of previously learned skills (e.g., from schooling). Barbey's theory has been supported by some preliminary evidence, but more is needed to confirm its validity (Girn et al., 2019). One potentially promising avenue of new empirical evidence is the emerging insight coming from longitudinal studies, which I describe in the next section.

Jumping through the Developmental Windowpane: Longitudinal Studies of Concurrent Structural Brain and Cognitive Changes from Childhood to Early Adulthood

Cross-sectional data provide incomplete information about the development of general intelligence

Any serious attempt at a mechanistic understanding of intelligence must involve consideration of developmental changes and trajectories. Longitudinal data, whereby participants undergo testing at two or more time points (also known as waves), are necessary to capture the dynamics of cognitive ability across the lifespan as cross-sectional data alone remain insufficient (see Raz and Lindenberger, 2011). Cross-sectional data can only be used to account for *between-subject differences* among individuals in a population at a single time point. For instance, if participant A scores higher than participants B and C on an IQ test on Monday, they are more intelligent according to that test, but only *relative* to their peers *on that day*. However, their score on Monday does not tell me how well they will score and compare with others throughout the week, or how they would develop over time. To answer these questions, I would need longitudinal data, which along with between-subject differences also permits the study of the *within-person changes* of participant A's IQ trajectory over time (i.e., how their score fluctuates from Monday to Tuesday to Wednesday, etc.). Furthermore, longitudinal data allow me to compare participant A's performance to themselves (i.e., rather than their peers), which provides *subject-specific* information about how participant A's intelligence changes throughout development (e.g., ranging from day-to-day to across the lifespan), as well as a comparison of between-subject differences in rates of change.

Echoing this point, one recent study (Schmiedek et al., 2020) has shown that a cross-sectional analysis of the *g* factor of cognitive ability failed to capture within-person changes in cognitive abilities over time. This highlights the need to integrate between-person differences (cross-sectional) and within-person changes (longitudinal) when studying cognitive abilities. Such an approach must be extended to include *concurrent* changes in brain structure and function to examine how brain and behaviour correlate with (i.e., baseline-baseline and baseline-second time point) and predict changes in each other (e.g., baseline-change (slope) in cognition and/or brain, and correlated change (cognitive slope and brain

slope)). In the next section, I will summarise the major findings from a recent review of longitudinal studies that *jointly* measure the relationship of cognitive ability to its neural correlates (particularly structural measures such as grey matter cortical volume and white matter fractional anisotropy) from childhood to early adulthood (Kievit and Simpson-Kent, 2021).[3] Doing so will further help elucidate the cognitive neuroscience of intelligence by also taking into account one of its central yet often neglected elements: (developmental) *time*.

The structural scaffolding of cognitive ability from childhood through early adulthood

The longitudinal cognitive neuroscience of intelligence from childhood to early adulthood is in its infancy (pun intended), with around 30 total studies to date (see Table 7.1 of Kievit and Simpson-Kent, 2021), many of which were published within the last five years. Despite varying sample sizes (minimum N for first wave = 33, maximum N for first wave = 2,091), cognitive measures (e.g., WISC, Mullen Scales of Early Learning, etc.), imaging metrics (e.g., cortical thickness, fractional anisotropy, etc.) and methodologies (e.g., simple correlations, latent change score models, etc.), three key interpretations stood out based on the aggregate findings from the studies: timing matters, methods matter, and convergent evidence in support of the *structural scaffolding* of intelligence (see Figure 7).

First, *timing* matters: the relationship between brain structure and cognitive ability shows both negative and positive associations, which might depend on age and ability (e.g., see Shaw et al., 2006), and/or sample size. This should not be surprising as both neural and cognitive performance change rapidly during development as the brain matures and children and adolescents learn through education. Therefore, the age distribution—for example, samples with a disproportionally large number of participants of a certain age group (e.g., pre-teenagers), and/or studies with a long interval between baseline and follow-up testing (i.e., longer than 4 years)—might not be suitable for capturing fine-grained changes in brain and behaviour.

Second, *methods* matter: due to methodological differences between most studies, direct comparisons among findings are difficult, if not impossible. However, this is only a challenge if one does not understand the differences (and similarities) between the

[3] For a recent review of correlated structural brain and cognitive changes in later life, see Oschwald et al., 2019.

quantitative methods used in longitudinal cognitive neuroscience. Different models entail different assumptions that can lead to different (and even seemingly contradictory) conclusions. Therefore, it is imperative that researchers understand not only the strengths but also the limitations of the statistical techniques they use to interpret data. Doing so allows for easier evaluations of theoretical frameworks across a variety of neurocognitive processes that encompass intelligence such as working memory and decision-making.

Finally, a consistent pattern emerged from the results of the studies (e.g., Ferrer, 2018; Wendelken et al., 2017) surveyed in the review: *current* brain structure (e.g., grey matter cortical thickness and white matter fractional anisotropy) but *not* function significantly predicted the *rate of change* in cognitive performance over time. In other words, individuals with 'better' brain structure such as greater white matter integrity were associated with larger gains (in children) or shallower declines (in older participants) than individuals with lower scores on structural brain imaging metrics. Although preliminary, this pattern suggests that the structural connectivity of the brain at a previous time point lays the foundations for later changes in cognition over time. This hypothesis is further tested in Chapter Four of this dissertation.

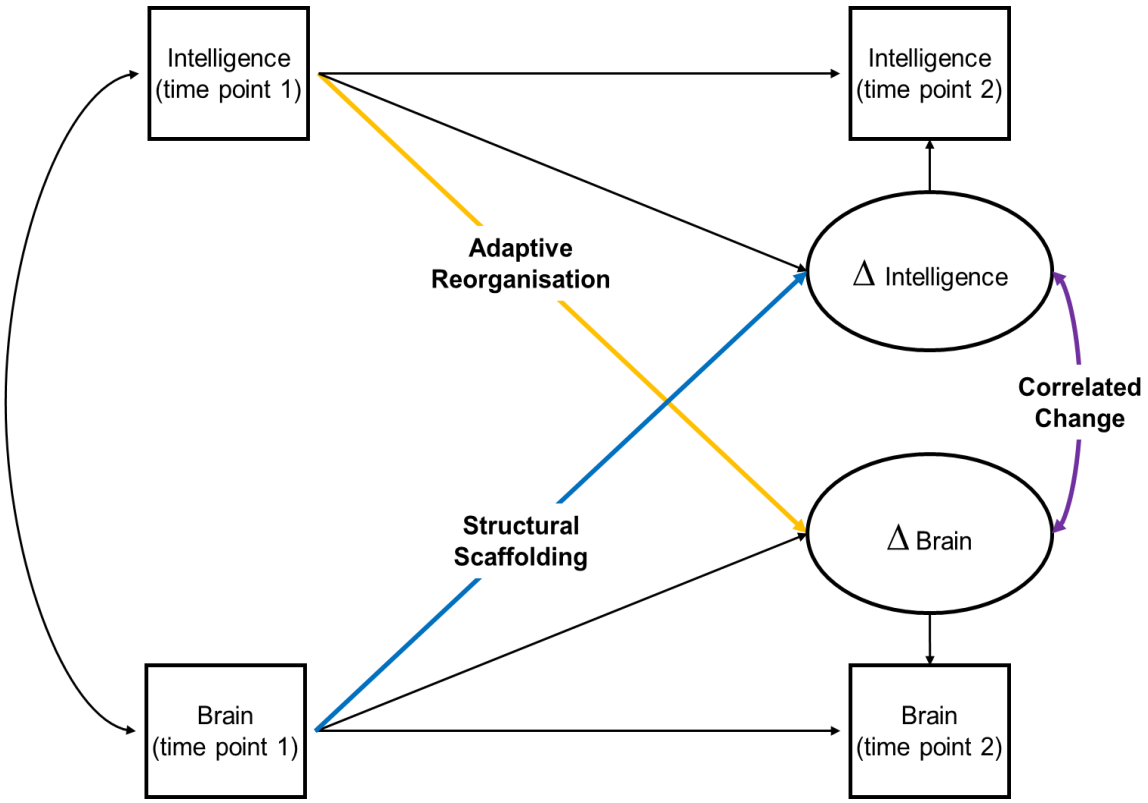


Figure 7. The 'structural scaffolding' model of human intelligence. Here, previous brain structure (time point 1) is thought to 'set the stage' from changes in intelligence (blue arrow). Alternatively (although not mutually exclusive of structural scaffolding), baseline cognition might predict changes in brain (structural) organisation over time (adaptive reorganisation, yellow arrow). Lastly, cognitive and brain changes might drive each other, or be driven by unmeasured other processes (correlated change, purple double-arrow).

Outline of the Thesis

This thesis attempts to advance understanding of human intelligence by investigating how human brain structure (i.e., cortical grey and white matter) and measures of cognitive ability such as working memory and decision-making co-develop with each other from childhood to early adulthood (ages 5 – 22 years). Towards this end, over the course of three empirical studies, I analysed neurocognitive data from three large cohorts (N range: 337 – 2072), combining insights from theory (i.e., *gc* and *gf* (Chapter Two), network neuroscience (Chapter Three), and brain-behaviour co-development (Chapter Four)) with rigorous and cutting-edge quantitative methods (i.e., structural equation modelling and network science). In the concluding remarks (Chapter Five), I argue that cognitive neuroscientists need to dedicate more time and effort towards building *coherent*, *rigorous*, and *testable* neurocognitive theories of intelligence—particularly through the concepts and tools of complex systems science.

Chapter Two

Neurocognitive Reorganisation between Crystallised Intelligence, Fluid Intelligence, and White Matter Integrity in Childhood and Adolescence

Chapter Two has been published (although appears in this thesis in a modified form to ensure consistent formatting and thesis coherence):

Simpson-Kent I. L., Fuhrmann D., Bathelt J., Achterberg J., Borgeest G. S., the CALM Team, Kievit R. A. (2020). Neurocognitive reorganization between crystallized intelligence, fluid intelligence and white matter microstructure in two age-heterogeneous developmental cohorts. *Developmental Cognitive Neuroscience, Special Issue: Flux 2018: Mechanisms of Learning & Plasticity*. doi: 10.1016/j.dcn.2019.100743

Although first-person pronouns are used throughout the chapter, this work is the result of a collaborative research project. My contributions to the above publication are:

1. I led conceptualisation and planning (aided by R. A. Kievit and D. Fuhrmann) about the scientific hypotheses, analysis methods, and interpretations of the project.
2. I performed all manuscript analyses (aided by R. A. Kievit and D. Fuhrmann).
3. I wrote the first full draft (with input from R. A. Kievit and D. Fuhrmann) of the manuscript and led the revisions and confirmation of the final version (aided by other co-authors) of the manuscript for publication in *Developmental Cognitive Neuroscience*.

Introduction

In Chapter One, I described the positive manifold and the theoretical foundation and predictive ability of general intelligence or *g* (Spearman, 1904). Moreover, I also discussed how Spearman's student, Raymond Cattell, proposed a division of Spearman's *g* factor into two separate yet related constructs, crystallised (*gc*) and fluid (*gf*) intelligence (Cattell, 1967). In this chapter, I use a novel statistical method, structural equation modelling trees, to study how the associations between *gc* and *gf*, and white matter (specifically fractional anisotropy) differ from childhood to adolescence (ages 5 – 18 years) in two cross-sectional, age-heterogeneous developmental samples (CALM and NKI-Rockland).

Fluid and crystallised intelligence have proven especially insightful regarding developmental changes in intelligence. For instance, current understanding of lifespan trajectories of *gc* and *gf* using cross-sectional (Horn and Cattell, 1967) and longitudinal (McArdle et al., 2000; Schaie, 1994) cohorts indicates that *gc* slowly improves until late age while *gf* increases into early adulthood before steadily decreasing. However, most of the literature on individual differences between *gc* and *gf* has focused on early to late adulthood. As a result, considerably less is known about the association between *gc* and *gf* in childhood and adolescence (but see Hülür et al., 2011).

There has, however, been a recent rise in interest in this topic in child and adolescent samples. For instance, research on age-related differentiation and its inverse, age dedifferentiation, in younger samples has greatly expanded since first being pioneered in the middle of the 20th century (Garrett, 1946). According to the *age differentiation hypothesis*, cognitive factors become less correlated (more differentiated) with increasing age. For example, the relationship (covariance) between *gc* and *gf* would decrease as children age into adolescence, suggesting that cognitive abilities increasingly specialise into adulthood. In contrast, the *age dedifferentiation hypothesis* predicts that cognitive abilities become more strongly related (less differentiated) throughout development. In this case, *gc* and *gf* covariance would increase between childhood and adolescence, potentially indicating a strengthening of the *g* factor across age. However, despite its increased attention in the literature, the age differentiation/dedifferentiation debate remains unsolved as evidence in support of both hypotheses has been found in child, adolescent, and adult samples (Bickley et al., 1995; de Mooij et al., 2018; Gignac, 2014; Hülür et al., 2011; Juan-Espinosa et al., 2000;

Tideman and Gustafsson, 2004). Together, this literature highlights the importance of a lifespan perspective on theories of cognitive development, as neither age differentiation nor dedifferentiation may be solely able to capture the dynamic changes that occur from childhood to adolescence and (late) adulthood (Hartung et al., 2018).

The introduction of non-invasive brain imaging technology has complemented conventional psychometric approaches by allowing for fine-grained probing of the neural bases of human cognition. A particular focus in developmental cognitive neuroscience has been the study of white matter using techniques such as diffusion-weighted imaging, which allows for the estimation of white matter microstructure (Wandell, 2016). Both cross-sectional and longitudinal research in children and adolescents using fractional anisotropy (FA), a commonly used estimate of white matter integrity, have consistently revealed strong correlations between FA and cognitive ability using tests of working memory, verbal and non-verbal performance (Koenis et al., 2015; Krogsrud et al., 2018; Peters et al., 2014; Tamnes et al., 2010; Urger et al., 2015) and even mathematical giftedness (Navas-Sánchez et al., 2014). However, interpretations of these studies are limited due to restricted cognitive batteries (e.g., small number of tests used) and a dearth of theory-driven statistical analyses (e.g., structural equation modelling).

For these reasons, several outstanding questions in the developmental cognitive neuroscience of intelligence remain: 1) Are the white matter substrates underlying intelligence in childhood and adolescence best understood as a single global factor, or do individual tracts provide specific contributions to *gc* and *gf*?, 2) If they are specific, are the tract contributions identical between *gc* and *gf*?, and 3) Does this brain-behaviour mapping differ across age (i.e., a neurocognitive interpretation of age differentiation/dedifferentiation or both)?

To examine these questions, I statistically tested four preregistered hypotheses:

1) *gc* and *gf* are separable constructs in childhood and adolescence. More specifically, the covariance among scores on cognitive tests are more adequately captured by the two-factor (*gc-gf*) model as opposed to a single-factor (i.e., *g*) model.
2) The covariance between *gc* and *gf* differs (decreases) across childhood and adolescence, in line with the age differentiation hypothesis.
3) White matter tracts make unique complementary contributions to *gc* and *gf*.
4) The contributions of these tracts to *gc* and *gf* differ (decrease) with age.

To address these questions, I examined the relationship between *gc* and *gf* in two large cross-sectional child and adolescent samples. The first is the Centre for Attention, Learning and Memory (CALM, see Holmes et al., 2019). This atypical sample, included in the preregistration (see https://aspredicted.org/5pz52.pdf), generally includes children with slightly lower cognitive abilities than age-matched controls (see Methods for more detail). To examine whether findings from CALM would generalise to other samples, I also conducted non-preregistered replication analyses on the Nathan Kline Institute (NKI)-Rockland Sample, a cohort with similar population demographics to the United States (e.g., race and socioeconomic status, see Table 1 of Nooner et al., 2012). All analyses were carried out using structural equation modelling (SEM), a multivariate statistical framework combining factor and path analysis to examine the extent to which causal hypotheses concerning latent (unobserved, e.g., *g*) and manifest (observed, e.g., cognitive tests scores) variables are in line with the observed data (Schreiber et al., 2006). Taken together, this chapter sought to investigate the relationship between measures of intelligence (i.e., *gc*, *gf* and working memory) and white matter connectivity (i.e., fractional anisotropy) in typically and atypically (struggling learners) developing children and adolescents.

Methods

Participants: The CALM and NKI-Rockland cohorts

For the CALM sample, I analysed the then most recent data release (N = 551, age range: 5 – 18 years) at the time of preregistration (January 2018). Participants were recruited based on referrals made for possible attention, memory, language, reading and/or mathematics problems (Holmes et al., 2019). Participants with or without formal clinical diagnosis were referred to CALM. Exclusion criteria included known significant and uncorrected problems in vision or hearing and a native language other than English. A subset of participants completed MRI scanning (N = 165, age range: 6 – 18 years).

Next, to assess the generalisability of the findings in CALM, I applied the same analyses to a non-preregistered subset of the data from the Nathan Kline Institute (NKI)-Rockland Sample (N cognitive data = 337, age range: 6 – 18 years; N neural data = 65, age range: 7 – 18 years). This multi-institutional initiative recruited a lifespan (aged between 6 and 85 years), community-ascertained sample (Nooner et al., 2012). This sample was chosen due to its representativeness (demographics resemble those of the United States population) and the fact that its cognitive battery assessments closely matched CALM.

Cognitive assessments: *gc*, *gf*, and working memory

All cognitive data from the CALM sample were collected on a one-to-one basis by an examiner in a dedicated child-friendly testing room. The test battery included a wide range of standardised assessments of cognition and learning (Holmes et al., 2019). Participants were given regular breaks throughout the session. Testing was divided into two sessions for participants who struggled to complete the assessments in one sitting. For analyses of the NKI-Rockland Sample cohort, I matched tasks used in CALM except for the Peabody Picture Vocabulary Test, Dot Matrix, and Mr. X, which were only available for CALM. For the NKI-Rockland Sample (Nooner et al., 2012), I included the N-Back task, which is not available in CALM, to help balance the number of working memory tasks between cohorts. In both samples, only raw scores obtained from assessments were included in analyses. Due to varying delays between recruitment and testing in the NKI-Rockland cohort, I only used cognitive test scores completed no later than six months (i.e., 180 days) after initial

recruitment. The cognitive tasks are further described in Table 1; the raw scores are depicted in Figure 8.

The working memory digit recall/span tasks, while measuring the same cognitive abilities, used different test batteries and scoring protocols. For CALM, working memory scores indicate the total number of correctly recalled digits across all trials while the NKI-Rockland scores were transformed into a span score. Due to this discrepancy (see References in Table 1 and Alloway et al., 2008 for statistical comparisons between the batteries), these tasks are plotted separately (see Figure 8).

Cognitive Domain	Task and Description	Mean (sd) [range]	Missing Data %	Reference
Crystallized Ability (*gc*)	**Peabody Picture Vocabulary Test (PPVT):** Participants were asked to choose the picture (out of four multiple-choice options) showing the meaning of a word spoken by an examiner.	**CALM:** 133.77 (31.68) [8, 215] **NKI-Rockland:** N/A	**CALM:** 1.09 **NKI-Rockland:** N/A	Dunn and Dunn, 2007
	Single Word Reading (SWR): Participants read aloud first a list of letters and then words that gradually increased in complexity. Correct responses required correctness and fluency.	**CALM:** 80.95 (24.35) [7, 130] **NKI-Rockland:** 104.47 (20.28) [35, 131]	**CALM:** 2.36 **NKI-Rockland:** 0	Wechsler, 2005
	Spelling (Spell): Participants spelled words with increasing difficulty one at a time that were spoken by an examiner.	**CALM:** 21.17 (8.68) [0, 48] **NKI-Rockland:** 33.57 (10.55) [4, 52]	**CALM:** 3.09 **NKI-Rockland:** 0	
	Numerical Operations (NO): Participants answered written mathematical problems that increased in difficulty.	**CALM:** 14.83 (7.46) [0, 48] **NKI-Rockland:** 27.95 (11.95) [4, 53]	**CALM:** 13.61 **NKI-Rockland:** 0	

Fluid Ability (*gf*)	**Matrix Reasoning (MR):** Participants saw sequences of partial matrices and selected the response option that best completed each matrix.	**CALM:** 10.88 (5.44) [0, 25] **NKI-Rockland:** 17.37 (5.19) [4, 27]	**CALM:** 0 **NKI-Rockland:** 0	Wechsler, 1999 Wechsler, 2011
Working Memory (WM)	**Digit Recall/Span (DR):** Participants recalled sequences of single digit numbers given in audio format.	**CALM:** 24.22 (5.32) [7, 43] **NKI-Rockland:** 5.97 (1.25) [3, 9]	**CALM:** 0.36 **NKI-Rockland:** 24.63	Alloway, 2007 Kaufman, 1975
	Backward Digit Recall/Span (BDR): Same as regular digit recall/span but in reversed order.	**CALM:** 9.2 (4.42) [0, 25] **NKI-Rockland:** 4.04 (1.40) [0, 8]	**CALM:** 1.63 **NKI-Rockland:** 24.63	
	Dot Matrix (Dot): Participants were shown the location of a red dot in a sequence of 4x4 matrices and had to recollect the location and order of these sequences.	**CALM:** 17.94 (5.49) [2, 35] **NKI-Rockland:** N/A	**CALM:** 0.18 **NKI-Rockland:** N/A	
	Mr. X (MRX): Participants remembered spatial locations of a ball held by a cartoon man rotated in one of seven positions.	**CALM:** 8.94 (4.90) [0, 30] **NKI-Rockland:** N/A	**CALM:** 0.91 **NKI-Rockland:** N/A	
	N-Back (NB): For 500 ms participants were presented letter sequences with a further 2000 ms to respond by pressing the computer spacebar. The task consisted of three separate conditions: 0-Back– participants pressed the spacebar whenever an "X" appeared; 1-Back– participants pressed the spacebar whenever the same letter was presented twice in a row; and, lastly, 2-Back– participants	**CALM:** N/A **NKI-Rockland:** 16.32 (4.22) [0, 20]	**CALM:** N/A **NKI-Rockland:** 20.47	Gur et al., 2010

	pressed the spacebar each time the letter presented matched the one shown two letters beforehand.			

Table 1. List, descriptions, and summary statistics (mean, standard deviation, range, and percentage of missing data) of cognitive assessments used in CALM and NKI-Rockland samples.

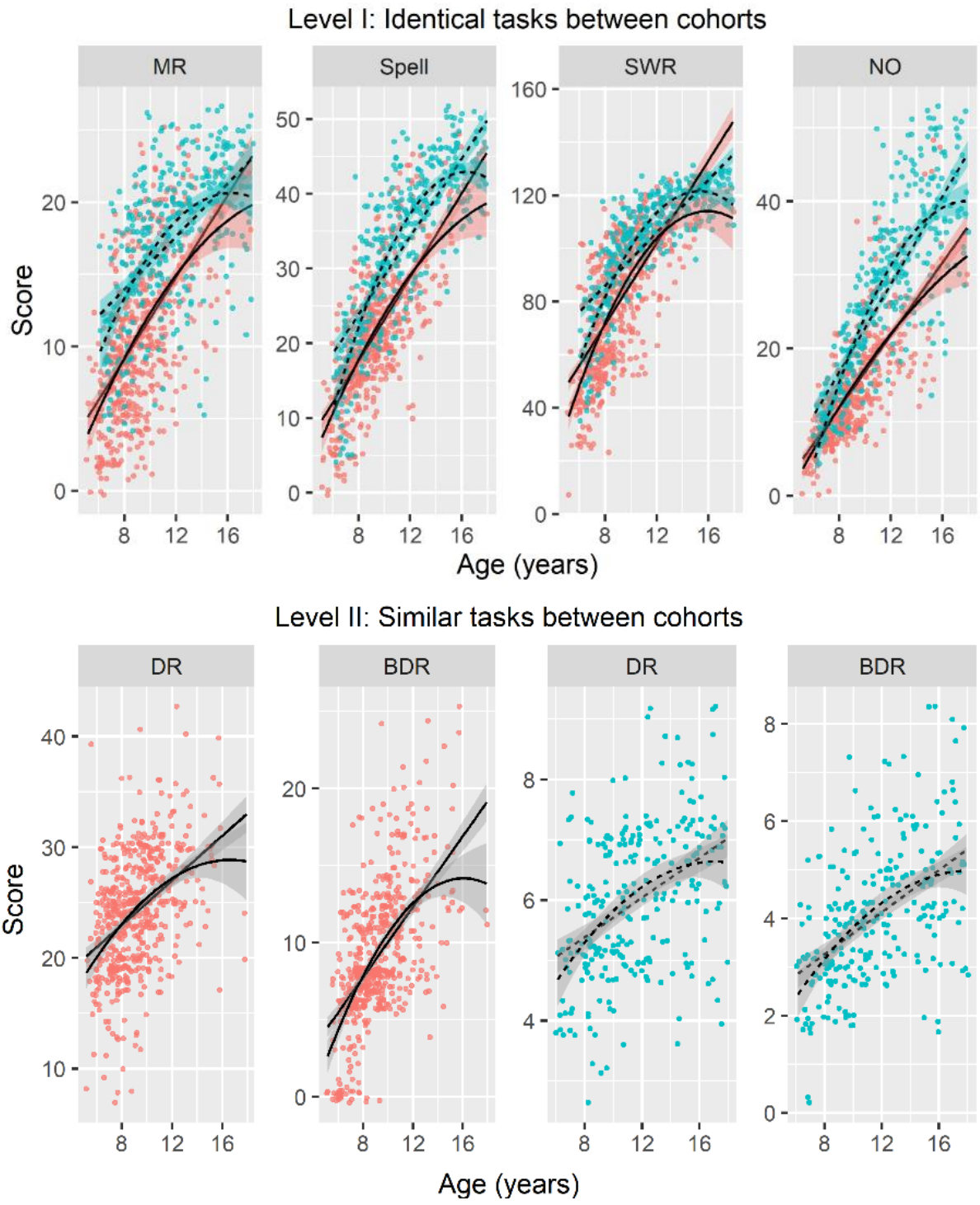


Figure 8 Top. Scatterplots of cognitive task scores across age for CALM and NKI-Rockland samples (Level I: tasks identical between cohorts; Level II: tasks similar between cohorts). Lines and shades reflect linear and polynomial fit and 95% confidence intervals, respectively. Solid lines: CALM. Dashed lines: NKI-Rockland. Abbreviations: Matrix Reasoning (MR), Spelling (Spell), Single Word Reading (SWR), Numerical Operations (NO), Forward Digit Recall/Span (DR), and Backward Digit Recall/Span (BDR).

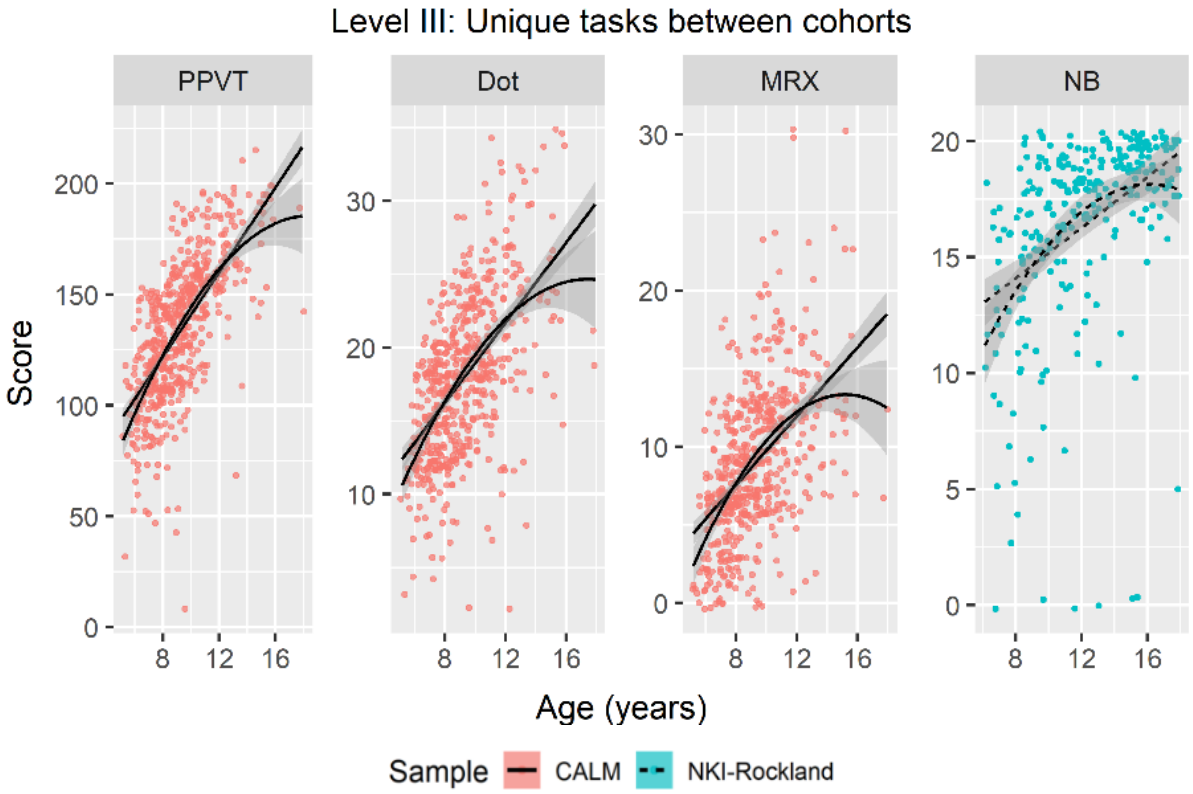


Figure 8 Bottom. Scatterplots of cognitive task scores across age for CALM and NKI-Rockland samples (Level III: tasks unique between cohorts). Lines and shades reflect linear and polynomial fit and 95% confidence intervals, respectively. Solid lines: CALM. Dashed lines: NKI-Rockland. Abbreviations: Peabody Picture Vocabulary Test (PPVT), Dot Matrix (Dot), Mr. X (MRX), and N-Back (NB).

Structural MRI measures: Fractional anisotropy (FA)

The CALM sample neuroimaging data were obtained at the MRC Cognition and Brain Sciences Unit, Cambridge, UK. Scans were acquired on the Siemens 3T Tim Trio system (Siemens Healthcare, Erlangen, Germany) via a 32-channel quadrature head coil. All T1-weighted volume scans were acquired using a whole brain coverage 3D magnetisation-prepared rapid acquisition gradient echo (MPRAGE) sequence with 1 millimetre (mm) isotropic image resolution with the following parameters: Repetition Time (TR) = 2250 milliseconds (ms); Echo Time (TE) = 3.02 ms; Inversion Time (TI) = 900 ms; flip angle = 9 degrees; voxel dimensions = 1 mm isotropic; GRAPPA acceleration factor = 2. Diffusion-Weighted Images (DWI) were acquired using a Diffusion Tensor Imaging (DTI) sequence with 64 diffusion gradient directions with a b-value of 1000 s/mm^2, plus one image acquired with a b-value of 0. Other relevant parameters include: TR = 8500 ms, TE = 90 ms, voxel dimensions = 2 mm isotropic.

The NKI-Rockland high-resolution 3D T1-weighted structural images were obtained using a Magnetisation Prepared Rapid Gradient Echo (MPRAGE) sequence with the following parameters: Repetition Time (TR) = 1900 ms; Echo Time (TE) = 2.52 ms; Inversion Time (TI) = 900 ms; flip angle = 9 degrees; voxel dimensions = 1 mm isotropic (see http://fcon_1000.projects.nitrc.org/indi/enhanced/NKI_MPRAGE.pdf for additional details). Diffusion-Weighted Images (DWI) were acquired with a Diffusion Tensor Imaging (DTI) sequence with 137 diffusion gradient directions with a b-value of 1500 s/mm^2. Other relevant parameters include: TR = 2400 ms, TE = 85 ms, voxel dimensions = 2 mm isotropic (see http://fcon_1000.projects.nitrc.org/indi/pro/eNKI_RS_TRT/DIff_137.pdf for additional details).

Note that part of the following pipeline for the white matter construction is identical to that described in Bathelt et al., 2019. Diffusion-weighted images were pre-processed to create a brain mask based on the b0-weighted image (FSL BET; Smith, 2002) and to correct for movement and eddy current-induced distortions (eddy; Graham et al., 2016). Subsequently, the diffusion tensor model was fitted and fractional anisotropy (FA) maps were calculated (dtifit). Images with a between-image displacement greater than 3 mm as indicated by FSL eddy were excluded from further analysis. All steps were carried out with FSL v5.0.9 and were implemented in a pipeline using NiPyPe v0.13.0 (Gorgolewski et al., 2011). To extract FA values for major white matter tracts, FA images were registered to the FMRIB58 FA template in MNI space using a sequence of rigid, affine, and symmetric diffeomorphic image registration (SyN) as implemented in ANTS v1.9 (Avants et al., 2008). Visual inspection indicated good image registration for all participants. Subsequently, binary masks from a probabilistic white matter atlas (threshold at > 50% probability) in the same space were applied to extract FA values for white matter tracts (see below).

Participant movement, particularly in developmental samples, can significantly affect the quality, and, hence, statistical analyses of MRI data. Therefore, several procedures were undertaken to ensure adequate MRI data quality and minimise potential biases due to subject movement. First, for the CALM sample, children were trained to lie still inside a realistic mock scanner prior to their actual scans. Secondly, for both samples, all T1-weighted images and FA maps were visually examined by a qualified researcher to remove low quality scans. Lastly, quality of the diffusion-weighted data was evaluated in both samples by calculating

the framewise displacement between subsequent volumes in the sequence. Only data with a maximum between-volume displacement below 3 mm were included in the analyses. All steps were carried out with FMRIB Software Library v5.0.9 and implemented in the pipeline using NiPyPe v0.13.0 (see https://nipype.readthedocs.io/en/latest/).

Selection of FA tracts was based on previous studies of associations between cognition (e.g., fluid intelligence) and white matter in developmental samples (de Mooij et al., 2018; Kievit et al., 2016). I used FA as a general summary metric of white matter microstructure as it cannot directly discern between specific cellular components (e.g., axonal diameter, myelin density, and water fraction). Mean FA was computed for 10 bilateral tracts as defined by the Johns Hopkins University DTI-based white matter tractography atlas (see Hua et al., 2008): anterior thalamic radiations (ATR), corticospinal tract (CST), cingulate gyrus (CING), cingulum [hippocampus] (CINGh), forceps major (FMaj), forceps minor (FMin), inferior fronto-occipital fasciculus (IFOF), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), and uncinate fasciculus (UNC). Figure 9 shows visualisations (top) and cross-sectional trends (bottom) of FA across the age range for both samples.

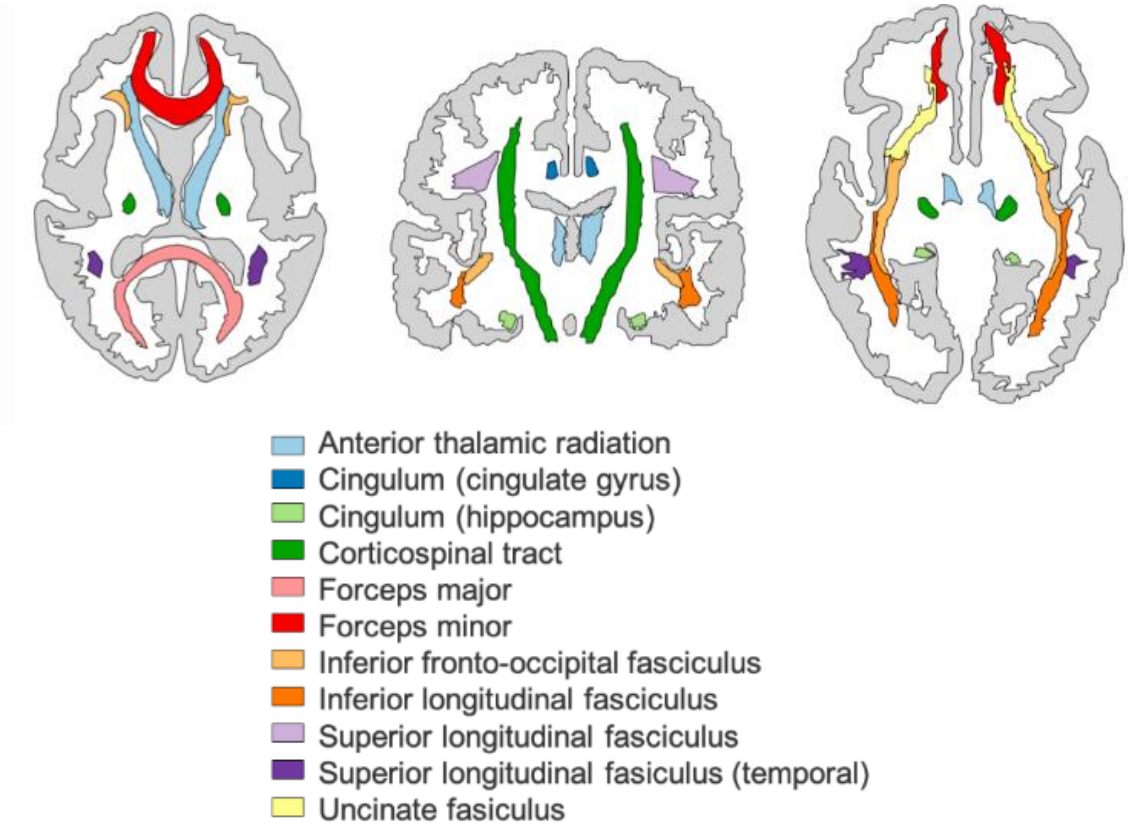


Figure 9. Top: White matter ROIs based on the John's Hopkin's University atlas (fractional anisotropy in Transverse (superior) plane (left), Coronal plane (middle), and Transverse (inferior) plane (right).

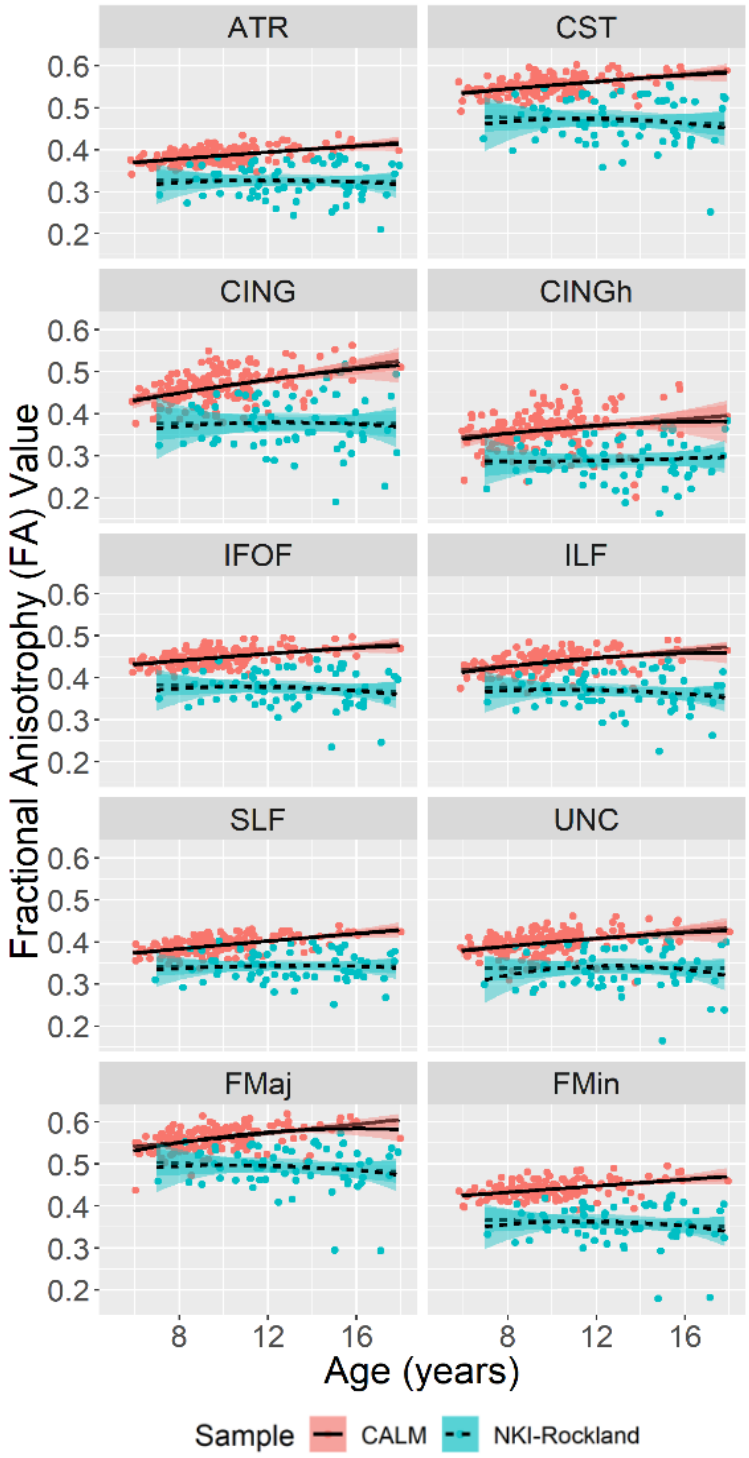


Figure 9. Bottom: Scatterplots of FA values for all white matter tracts across age for CALM and NKI- Rockland samples. Lines and shades reflect linear and polynomial fit and 95% confidence intervals, respectively. Solid lines: CALM. Dashed lines: NKI-Rockland. Abbreviations: anterior thalamic radiations (ATR), corticospinal tract (CST), cingulate gyrus (CING), cingulum [hippocampus] (CINGh), inferior fronto-occipital fasciculus (IFOF), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), uncinate fasciculus (UNC), forceps major (FMaj), and forceps minor (FMin).

Statistical analyses: Structural equation modelling (SEM) and SEM trees

I used structural equation modelling (SEM), a multivariate approach that combines latent variables and path modelling to test causal hypotheses (Schreiber et al., 2006) as well as SEM trees, which combine SEM and decision tree paradigms to simultaneously permit exploratory and confirmatory data analysis (Brandmaier et al., 2013). Analyses were performed using the lavaan package version 0.5-22 (Rosseel, 2012) in R (R Core Team, 2020) and versions 2.9.9 and 0.9.12 of the R packages OpenMx (Boker et al., 2011) and semtree (Brandmaier et al., 2013), respectively. To account for missing data and deviations from multivariate normality, I used robust full information maximum likelihood estimator (FIML) with a Yuan-Bentler scaled test statistic (MLR) and robust standard errors (Rosseel, 2012). I evaluated overall model fit via the (Satorra-Bentler scaled) chi-squared test, the comparative fit index (CFI), the standardised root mean squared residuals (SRMR), and the root mean square error of approximation (RMSEA) with its confidence interval (Schermelleh-Engel et al., 2003). Assessment of model fit was defined as: CFI (acceptable fit 0.95 – 0.97, good fit > 0.97), SRMR (acceptable fit 0.05 – 0.10, good fit < 0.05), and RMSEA (acceptable fit 0.05 – 0.08, good fit < 0.05).

To determine whether *gc* and *gf* were separable constructs, I compared a two-factor (*gc*-*gf*) model to a single-factor (*g*) model. To investigate if the covariance between *gc* and *gf* differed across ages, I conducted multiple group comparisons between younger and older participants based on median splits (CALM split at 8.91 years yielding N = 279 young and N = 272 old; NKI-Rockland split at 11.38 years into N = 169 young and N = 168 old). Doing so inevitably led to slightly unbalanced numbers of participants with white matter data (CALM: N = 60 young and N = 105 old; NKI-Rockland: N = 19 young and N = 46 old). To test measurement invariance across age groups (Putnick and Bornstein, 2016), I fit multigroup models (French and Finch, 2008), constraining key parameters across groups. Model comparisons and deviations from measurement invariance were determined using the likelihood ratio test and Akaike information criterion (AIC, see Bozdogan, 1987).

To examine whether white matter tracts made unique contributions to the latent variables (i.e., *g*, *gc*, and *gf*) I fit Multiple Indicator, Multiple Cause (MIMIC) models (Jöreskog and Goldberger, 1975; Kievit et al., 2012). MIMIC models enable neural variables (e.g., individual white matter tracts) to be *simultaneously* regressed onto cognitive latent constructs

such as *gc* and *gf*. Lastly, I conducted a SEM tree analysis, a method that combines the confirmatory nature of SEM with the exploratory framework of decision trees (Brandmaier et al., 2013). SEM trees hierarchically and recursively partition datasets based on a covariate (in this case age). This creates data-driven age-groups which show differences in one or more paths of interest. The advantage of SEM trees is that they do not require a-priori decisions as to where potential categorical boundaries between age groups may lie (as was the case in the median split analysis). SEM trees also do not require a-priori knowledge as to the shape of developmental trajectories (as is usually the case when using age as a continuous covariate). Using this technique therefore allowed for the examination of: 1) the robustness of findings based on the median age split, and 2) whether white matter contributions differed across age groups of younger and older participants in a data-driven way (Hypothesis 4). Therefore, for the CALM and NKI-Rockland SEM tree analyses, age was used as a continuous covariate with a significance alpha level of .001 for each node split. Finally, I used Bonferroni-correction to correct for multiple comparisons, and the semtree cross-validation scheme, which "partitions the data for maximizing splits on each variable, then comparing maximum splits across each variable on the rest of the data" (for more details, see https://cran.r-project.org/web/packages/semtree/semtree.pdf).

Results

Covariance among cognitive abilities cannot be captured by a single factor

In accordance with the preregistered analysis plan, I first describe model fit for the measurement models of the cognitive data only. First, I tested hypothesis 1: that *gc* and *gf* are separable constructs in childhood and adolescence. More specifically, I tested the hypothesis that the covariance among scores on cognitive tests would be better captured by a two-factor (*gc-gf*) model than a single-factor (i.e., *g*) model. In support of this prediction, the single-factor model fit the data poorly: $\chi^2(27) = 317.695$, $p < .001$, CFI = .908, SRMR = .040, RMSEA = .146 [.132 .161], Yuan-Bentler scaling factor = 1.090, suggesting that cognitive performance was not well represented by a single factor. The two-factor (*gc-gf*) model also displayed poor model fit ($\chi^2(24) = 196.348$, $p < .001$, CFI = .946, SRMR = .046, RMSEA = .119 [.104 .135], Yuan-Bentler scaling factor = 1.087), although it fit significantly better ($\chi^2\Delta = 119.41$, $df\Delta = 3$, $AIC\Delta = 127$, $p < .001$) than the single-factor model.

To investigate the source of poor fit, I examined modification indices (Schermelleh-Engel et al., 2003), which quantify the expected improvement in model fit if a parameter is freed. Modification indices suggested that the Peabody Picture Vocabulary Test had a very strong cross-loading onto the fluid intelligence latent factor. The Peabody Picture Vocabulary Test (PPVT), often considered a crystallised measure in adult populations, asks participants to choose the picture (out of four multiple-choice options) corresponding to the meaning of the word spoken by an examiner. Including a cross-loading between *gf* and the PPVT drastically improved goodness of fit ($\chi^2\Delta = 67.52$, $df\Delta = 1$, $AIC\Delta = 100$, $p < .001$) to adequate ($\chi^2(23) = 104.533$, $p < .001$, CFI = .975, SRMR = .025, RMSEA = .083 [.067 .099], Yuan-Bentler scaling factor = 1.069). A likely explanation of this result is that such tasks may draw considerably more on executive, *gf*-like abilities in younger, lower ability samples. For a more thorough investigation of the loading of PPVT across development, see **Supplementary Material** section **Is the Peabody Picture Vocabulary Test a measure of fluid ability?**. Notably, fitting the PPVT as a *solely* fluid task (i.e., removing it as a measurement of *gc* entirely) did not significantly decrease model fit ($\chi^2\Delta = 2.058$, $df\Delta = 1$, $AIC\Delta = 1$, $p = .152$). Therefore, I decided to proceed with the more parsimonious PPVT *gf*-only model ($\chi^2(24) = 106.382$, $p < .001$, CFI = .972, SRMR = .025, RMSEA = .082 [.066 .098], Yuan-Bentler scaling factor = 1.073).

Next, I examined whether the single- or two-factor model fit best in the NKI-Rockland sample. The single-factor model fit the data adequately ($\chi^2(14) = 41.329$, $p < .001$, CFI = .983, SRMR = .029, RMSEA = .075 [.049 .102], Yuan-Bentler scaling factor = .965). Still, the two-factor model showed considerably better fit ($\chi^2(12) = 19.732$, $p = .072$, CFI = .995, SRMR = .018, RMSEA = .043 [.000 .075], Yuan-Bentler scaling factor = .956) compared to the single-factor model ($\chi^2\Delta = 20.661$, $df\Delta = 2$, $AIC\Delta = 17$, $p < .001$). It should be noted that, given the differences in tasks measured between the samples, *gf* and working memory were assumed to be measurements of the same latent factor, rather than separable factors. A similar competing model where *gf* and working memory were modelled as separate constructs with working memory loaded onto *gf*, similarly to the best-fitting model for the CALM sample (see Figure 10), showed comparable model fit and converging conclusions with further analyses. Overall, these findings suggested, that for both the CALM and NKI-Rockland samples, a two-factor model with separate *gc* and *gf* factors provided a better account of individual differences in intelligence than a single-factor model. Note that, for both CALM and NKI-Rockland, these final models include a latent variable called 'gc Verbal' that is loaded onto the more fundamental *gc* factor.

Evidence of age differentiation between crystallised and fluid ability

I investigated the relationship between *gc* and *gf* in development to see whether I could observe evidence for age differentiation as predicted by hypothesis 2. Age differentiation (e.g., Hülür et al., 2011) would predict decreasing covariance between *gc* and *gf* from childhood to adolescence. I fit a multigroup confirmatory factor analysis to assess fit on the younger (N = 279) and older (N = 272) participants from the CALM sample. The model had acceptable fit ($\chi^2(48) = 142.214$, $p < .001$, CFI = .960, SRMR = .037, RMSEA = .085 [.069 .102], Yuan-Bentler scaling factor = 1.019). However, a likelihood ratio test, showed that model fit did not decrease significantly when imposing equal covariance between *gc* and *gf* in the younger and older participant subgroups ($\chi^2\Delta = 0.323$, $df\Delta = 1$ $AIC\Delta = 2$, $p = .57$). This suggested no evidence for age differentiation in the CALM sample. However, the lack of association could be due to limitations of using median splits to investigate age differences when independent (or latent in this case) variables are correlated (Iacobucci et al., 2015). For instance, if the age range of differences in behavioural associations between *gc* and *gf* lies elsewhere, the median split may

not be sensitive enough to detect it. To test this explicitly, I next fit SEM trees (Brandmaier et al., 2013) to the cognitive data.

I estimated SEM trees in the CALM sample by specifying the cognitive model with age as a continuous covariate. I observed a SEM tree split at age 9.12 years, yielding two groups (younger participants = 290, older participants = 261). This split was accompanied by a decrease in the unstandardised parameter estimate between *gc* and *gf* (from .64 to .59, see Table 3), providing support for age differentiation using a more exploratory approach (SEM tree: 9.12 versus median split: 8.91). When fitting the two-factor model before and after the SEM tree age split, I found that the correlation between *gc* and *gf* increased slightly (from .90 to .92), which supports age dedifferentiation.

Next, as in the CALM cohort, I fit a multigroup model with younger (N = 169) and older (N = 168) age groups in the NKI-Rockland Sample, which produced good fit ($\chi^2(24)$ = 33.736, p = .089, CFI = .991, SRMR = .035, RMSEA = .047 [.000 .081], Yuan-Bentler scaling factor = .916). In contrast to CALM, imposing equality constraints on the covariance between *gc* and *gf* across age groups preferred the freely-estimated model ($\chi^2\Delta$ = 61.244, dfΔ = 1 AICΔ = 46, p < .001) and revealed a lower *gc-gf* correlation for the older (.811) compared to the younger participants (1.008). This suggested evidence for age differentiation in the NKI-Rockland Sample using multigroup models.

In contrast to the multigroup model outcome, the NKI-Rockland SEM tree model under identical specifications as in CALM failed to produce an age split. A possible explanation is that, to penalise for multiple testing, I relied on Bonferroni-corrected alpha thresholds for the SEM tree. If, as seems to be the case here, the true split lies (almost) exactly on the median split, then the SEM tree will have slightly less power than conventional multigroup models, as the SEM tree likelihood ratio test is penalised for the number of tests (splits). These differences between analyses methods suggested that the age differentiation observed here is likely modest in size. Taken together, I interpret these findings as evidence for a small, age-specific but suggestive decrease in *gc-gf* covariance in both cohorts, which is compatible with age differentiation such that, for younger participants, *gc* and *gf* factors are almost indistinguishable, whereas for older participants a clearer separation emerges.

Violation of metric invariance suggests differences in relationships among cognitive abilities in childhood and adolescence

I more closely assessed age-related differences in cognitive architecture (i.e., factor loadings) by examining metric invariance (Putnick and Bornstein, 2016). Testing this in the CALM sample as a two-group model by imposing equality constraints on the factor loadings (fully constrained) showed that the freely-estimated model (no factor loading constraints) outperformed the fully-constrained model ($\chi^2\Delta$ = 107.05, dfΔ = 7, AICΔ = 82, p < .001), indicating that metric invariance was violated. This violation of metric invariance suggested that the relationship between the cognitive tests and latent variables was different in the two age groups. Closer inspection suggested that the differences in loadings were not uniform, but rather showed a more complex pattern of age-related differences (see Table 2 for more details). Some of the most pronounced differences include an increase of the loading of Matrix Reasoning onto *gf* as well as increased loading of Digit Recall and Dot Matrix onto working memory across age groups.

Similarly, in the NKI-Rockland cohort, the freely-estimated model outperformed the constrained model ($\chi^2\Delta$ = 41.111, dfΔ = 5, AICΔ = 33, p < .001), indicating that metric invariance was again violated as in CALM. This suggests that the relationship between the cognitive tests and the latent factors differed across age groups. The pattern of factor loadings differed in some respects from CALM. For example, the loading of the N-back task onto *gf* showed the largest difference across age groups in the NKI-Rockland sample. However, as CALM did not include the N-back task, this finding cannot directly be interpreted as a difference between the cohorts. For detailed comparisons among factor loadings between age groups in both samples, refer to Table 2. The overall pattern in both samples suggested small and varied differences in the relationship between the latent factors and observed scores. A plausible explanation is that the same task draws on a different balance of skills as children differ in age and ability. Therefore, these findings concerning the latent factors should be interpreted in this light as it seems likely that in addition to age differentiation (and possibly dedifferentiation) effects, the nature of the factors also differed slightly across the age range studied here.

Relationship	CALM		NKI-Rockland	
	Young	Old	Young	Old
gc↔gf(WM)	0.89 40.23 (5.40) [29.65, 50.82]	0.93 44.27 (4.50) [35.45, 53.09]	1.01 64.49 (8.64) [47.56, 81.43]	0.81 12.27 (2.68) [7.02, 17.52]
gf→WM	0.96 1.06 (.19) [.69, 1.44]	0.9 .79 (.09) [.61, .97]	NA	NA
gf(WM)→MR	0.59 1.00 (NA) [1.00, 1.00]	0.74 1.00 (NA) [1.00, 1.00]	0.69 1.00 (NA) [1.00, 1.00]	0.6 1.00 (NA) [1.00, 1.00]
gf→PPVT	0.75 7.49 (.84) [5.84, 9.14]	0.76 5.45 (.43) [4.60, 6.30]	NA	NA
gf(WM)→DR	0.56 1.00 (NA) [1.00, 1.00]	0.68 1.00 (NA) [1.00, 1.00]	0.38 .12 (.03) [.07, .17]	0.54 .27 (.07) [.13, .40]
gf(WM)→BDR	0.76 1.01 (.12) [.77, 1.26]	0.79 .94 (.09) [.76, 1.12]	0.5 .16 (.03) [.10, .22]	0.53 .30 (.08) [.13, .40]
gfWM→NB	NA	NA	0.55 .67 (.10) [.48, .87]	0.35 .54 (.14) [.27, .81]
WM→Dot	0.58 .87 (.12) [.63, 1.10]	0.67 1.06 (.12) [.82, 1.30]	NA	NA
WM→MRX	0.59 .80 (.11) [.57, 1.02]	0.56 .82 (.13) [.56, 1.08]	NA	NA
gc→gcV	0.89 1.00 (NA) [1.00, 1.00]	0.79 1.00 (NA) [1.00, 1.00]	0.96 1.00 (NA) [1.00, 1.00]	0.87 1.00 (NA) [1.00, 1.00]
gc→NO	0.87 .19 (.01) [.17, .22]	0.87 .54 (.06) [.43, .65]	0.9 .42 (.03) [.36, .49]	0.76 1.08 (.20) [.69, 1.48]
gcV→SWR	0.94 1.00 (NA)	0.91 1.00 (NA)	0.93 1.00 (NA)	0.89 1.00 (NA)

	[1.00, 1.00]	[1.00, 1.00]	[1.00, 1.00]	[1.00, 1.00]
	0.87	0.91	0.97	0.88
gcV➔Spell	.28 (.02)	.46 (.03)	.48 (.02)	.71 (.06)
	[.25, .31]	[.40, .51]	[.44, .52]	[.60, .83]

Table 2. First row: standardised path estimates for cognitive assessments in CALM and NKI-Rockland samples. Second row: raw path estimates with standard errors (parentheses). Third row: 95% confidence intervals [brackets]. NA = not applicable. Note that age groups were determined according to the median split (CALM: 8.91 years, NKI-Rockland: 11.38 years).

The neural architecture of *gc* and *gf* indicates unique contributions of multiple white matter tracts to cognitive ability

I next focused on the white matter regression coefficients to inspect the neural underpinnings of *gc* and *gf*. In line with hypothesis 3, I wanted to explore whether individual white matter tracts made independent contributions to *gc* and *gf*. First, I examined whether a single-factor model could account for the covariance in white matter microstructure across the ten tracts (Figure 9). If so, then scores on such a latent factor would represent a parsimonious summary for neural integrity. However, this model showed poor fit ($\chi^2(35)$ = 124.810, p < .001, CFI = .938, SRMR = .039, RMSEA = .132 [.107 .157], Yuan-Bentler scaling factor = 1.114), suggesting that white matter integrity cannot be summarised by a single factor. To examine whether the white matter tracts showed specific and complementary associations with cognitive performance, I fit a MIMIC model in the CALM sample using the 10 white matter tracts mentioned above (see Figure 9). Doing so, I observed that 5 out of the 10 tracts showed significant relations with *gc* and/or *gf* (Figure 10). Specifically, the anterior thalamic radiations, forceps major, and forceps minor had moderate to strong associations with *gc* with similar relations seen for *gf* for the superior longitudinal fasciculus, forceps major, and the cingulate gyrus. Interestingly, the forceps minor exhibited a negative association with *gf*. This could be due to modelling several highly correlated paths simultaneously since this relationship was not found when only the forceps minor was modelled onto *gc* (standardised estimate = .426) and *gf* (standardized estimate = .386), see Tu et al., 2008. Together, individual differences in white matter microstructure explained 32.9% in crystallised and 33.6% in fluid ability.

As in the CALM sample, the single-factor white matter model produced poor fit (χ^2 (35) = 131.637, p < .001, CFI = .924, SRMR = .023, RMSEA = .201 [.165 .238], Yuan-Bentler scaling factor = .950) in the NKI-Rockland Sample. Therefore, I fit a multi-tract MIMIC model. The superior longitudinal fasciculus emerged as the only tract to significantly load onto *gc* or *gf* (Figure 10). This result was likely due to lower power associated with a small subset of individuals with white matter data (see Discussion for further investigation). In NKI-Rockland, the same set of tracts explained 29.7% and 26.7% of the variance in *gc* and *gf*, respectively. Together, these findings demonstrated generally similar associations (in terms of variance explained) between white matter microstructure and cognitive abilities in the CALM and NKI-Rockland samples. Therefore, it seems to be the case that, in both typically and atypically (struggling learners) developing children and adolescents, individual white matter tracts make distinct contributions to crystallised and fluid ability, as more than one tract explains variance in the outcomes (*gc* and *gf*).

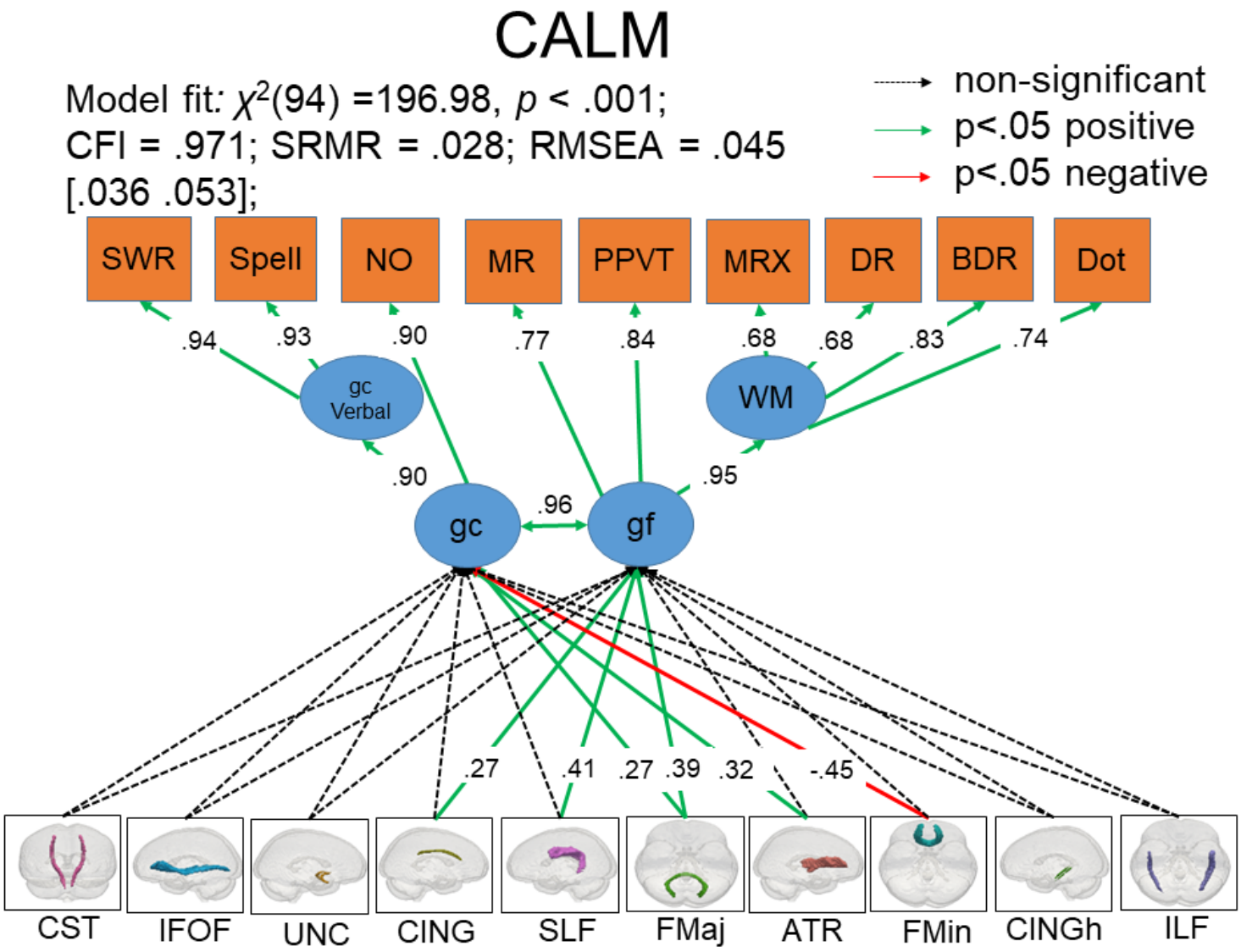


Figure 10 Top. MIMIC model displaying standardised parameter estimates and regression coefficients for all cognitive measures and white matter tracts for complete CALM sample. Dotted, green, and red arrows indicate nonsignificant (> .05), positively significant, and negatively significant path estimates, respectively.

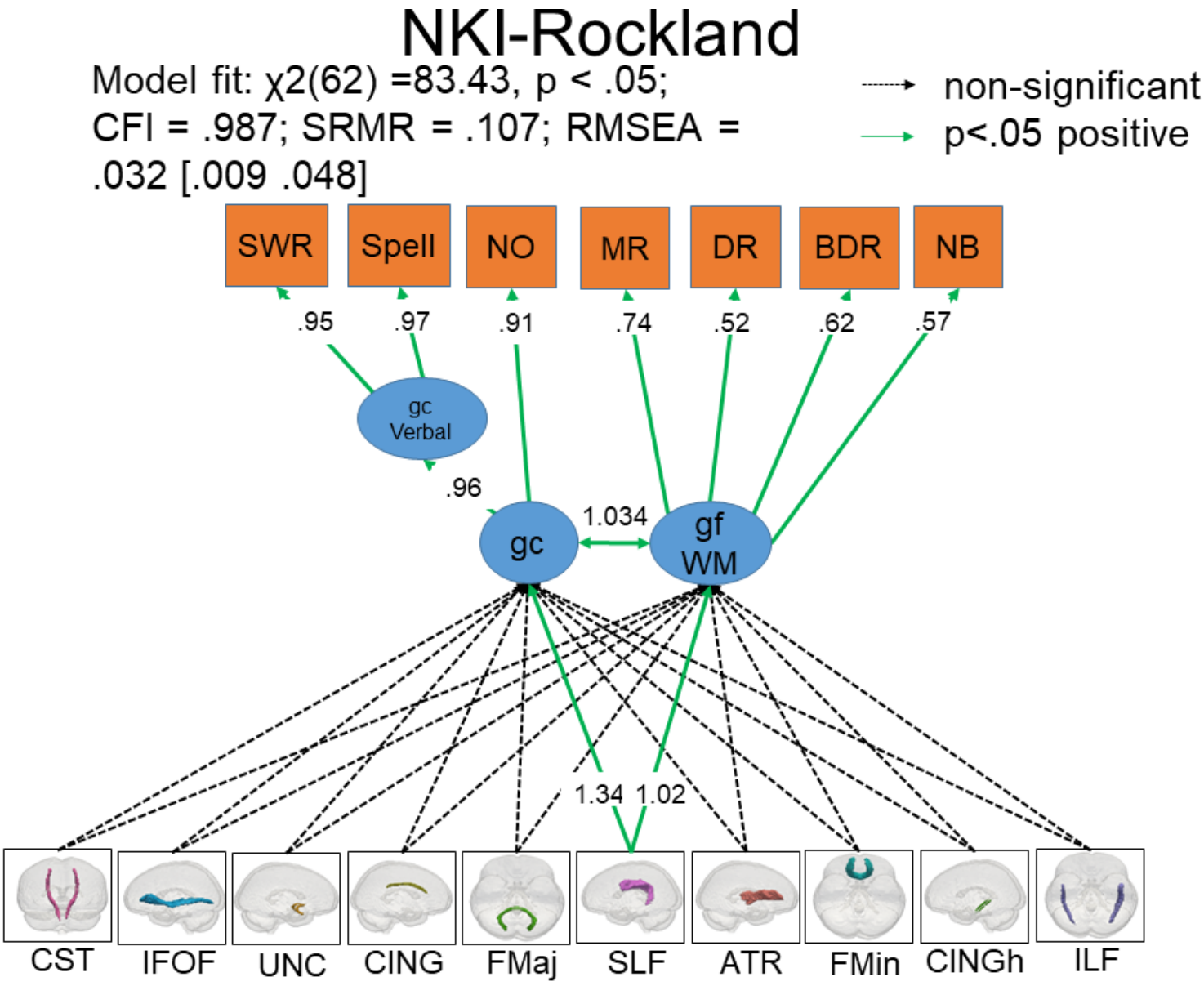


Figure 10 Bottom. MIMIC model displaying standardised parameter estimates and regression coefficients for all cognitive measures and white matter tracts for complete NKI-Rockland Sample. Dotted and green arrows indicate nonsignificant (> .05) and positively significant path estimates, respectively. Note standardised estimate exceeding 1 is likely the consequence of highly-correlated factors (Jöreskog, 1999).

Support for neurocognitive reorganisation of crystallised and fluid ability in childhood and adolescence

Lastly, to address the fourth and final hypothesis, I examined whether brain-behaviour associations differed across the developmental age range. I hypothesised that the relationship between the white matter tracts and cognitive abilities would decrease across the age range, in support of the differentiation hypothesis, inclusive of neural associations. Using

a multigroup model, I compared the strength of brain-behaviour relationships between younger and older participants to test whether white matter contributions to *gc* and *gf* differed in development. Contrary to this prediction, I observed that, in the CALM sample, a freely estimated model, where the brain-behaviour relationships were allowed to vary across age groups, did not outperform the constrained model ($\chi^2\Delta = 12.16$, $df\Delta = 10$, $AIC\Delta = 9$, $p = .27$). This suggested that the contributions of white matter tracts did not vary significantly between age groups when examined using multigroup models.

As before with the cognitive data, next I estimated a SEM tree model from the CALM MIMIC model. In contrast to the multigroup model, I observed that multiple white matter tracts did differ in their associations with *gc* and/or *gf*. These differences manifested in different ways for *gc* and *gf*. For example, the correlations between the cingulum, superior longitudinal fasciculus, and forceps major and *gf* decreased with increasing age, in line with age differentiation. On the other hand, the forceps major, forceps minor and anterior thalamic radiations demonstrated a more complicated pattern with each tract displaying two age splits. For the first split (around age 8), the regression strength decreased before spiking again around age 11 (Table 3). Given that all first splits showed a decrease between white matter and cognition, and all second splits revealed an increase compared to the first, this suggests a non-monotonic pattern of brain-behaviour reorganisation that cannot be fully captured by age differentiation or dedifferentiation (Hartung et al., 2018) but may be in line with theories such as Interactive Specialization (Johnson, 2011), which provides a range of mechanisms which may induce age-varying brain-behaviour strengths. One hypothesis that has previously been offered that may (partially) explain the nature of the age-varying associations between white matter and cognitive performance is the onset of puberty (Fuhrmann et al., 2020) and the associated hormonal changes. Previous work has shown that pubertal processes, including differences and changes in hormones such as testosterone, affect diffusion measures in ways that cannot be explained away by (only) age (Menzies et al., 2015). More work in large samples such as ABCD (Volkow et al., 2018), ideally including longitudinal changes in hormone levels, is needed to establish the robustness of this explanation.

Lastly, I performed the same multigroup analysis for the NKI-Rockland MIMIC model, but it failed to converge or produce an age split, likely due to sparsity of the neural data. Therefore, this analysis could not be used to replicate the cut-off age used for multigroup

analyses (11.38 years) based on the median split. Further inspection with SEM trees of the only significantly associated tract, the superior longitudinal fasciculus, revealed the same trend for *gc* and *gf* with decreased correlations with increasing age (see Table 3). Overall, the findings suggest the need for a neurocognitive account of age differentiation-dedifferentiation/reorganisation from childhood into adolescence.

Parameter	CALM					NKI-Rockland				
	Estimate before split	Age of 1st split	Estimate after split	Age of 2nd split	Estimate after split	Estimate before split	Age of 1st split	Estimate after split	Age of 2nd split	Estimate after split
gc←→gf(WM)	.64 (.01)	9.12	.59 (.01)	NS	NS	.96 (.02)	NS	NS	NS	NS
gf(WM)←SLF	.38 (.05)	7.38	.29 (.03)	NS	NS	.35 (.09)	13.16	.21 (.09)	NS	NS
gc←SLF	NA	NA	NA	NA	NA	.91 (.09)	9.85	.69 (.06)	NS	NS
gf←FMaj	.38 (.04)	7.38	.26 (.03)	NS	NS	NA	NA	NA	NA	NA
gc←FMaj	.24 (.04)	8.29	.04 (.05)	10.79	.42 (.05)	NA	NA	NA	NA	NA
gf←CING	.29 (.05)	7.38	.18 (.03)	NS	NS	NA	NA	NA	NA	NA
gc←ATR	.30 (.05)	7.62	.13 (.05)	10.79	.37 (.05)	NA	NA	NA	NA	NA
gc←FMin	-.34 (.06)	7.62	-.52 (.04)	10.79	-.25 (.05)	NA	NA	NA	NA	NA

Table 3. SEM tree Results for CALM and NKI-Rockland samples. Note: values listed represent unstandardised estimates and standard errors (parentheses). NS = no SEM tree split, NA = not applicable.

Discussion

Summary and interpretation of findings

In this chapter, I examined the cognitive architecture as well as the white matter substrates of fluid and crystallised intelligence in children and adolescents in two developmental samples (CALM and NKI-Rockland). Analyses in both samples indicated that individual differences in intelligence were better captured by two separate but highly correlated factors (*gc* and *gf*) of cognitive ability as opposed to a single global factor (*g*). Further analysis suggested that the covariance between these factors decreased slightly from childhood to adolescence, in line with the age differentiation hypothesis of cognitive abilities (Garrett, 1946; Hülür et al., 2011).

I observed multiple, partially independent contributions of specific tracts to individual differences in *gc* and *gf* (Figure 10). The clearest associations were observed for the anterior thalamic radiations, cingulum, forceps major, forceps minor, and superior longitudinal fasciculus, all of which have been implicated to play a role in cognitive functioning in childhood and adolescence (Krogsrud et al., 2018; Navas-Sánchez et al., 2014; Peters et al., 2014; Tamnes et al., 2010; Urger et al., 2015; Vollmer et al., 2017). However, except for the superior longitudinal fasciculus, these tracts were not significant in the NKI-Rockland Sample. A possible explanation for this is the difference in imaging sample size between the cohorts (N = 165 in the CALM sample vs N = 65 in the NKI-Rockland Sample). This difference implies sizeable differences in power (73.4% in CALM versus 36.2% in NKI-Rockland, assuming a standardised effect size of 0.2) to identify weaker individual pathways.

The most consistent association, observed in both samples, was between the superior longitudinal fasciculus, a region known to be important for language and cognition, which significantly contributed to cognitive ability in both CALM (*gf* only) and NKI-Rockland (*gc* and *gf*). The superior longitudinal fasciculus is a long myelinated bidirectional association fibre pathway that runs from anterior to posterior cortical regions and through the major lobes of each hemisphere (Kamali et al., 2014), and has been associated with memory, attention, language, and executive function in childhood and adolescence in both healthy and atypical populations (Frye et al., 2010; Urger et al., 2015). Therefore, given its widespread links throughout the brain, which include temporal and fronto-parietal regions, it is no surprise that it was found to be significantly related to both *gc* and *gf* in the two samples.

Together, these results are in line with previous research relating fractional anisotropy (FA) and cognitive ability. For instance, Peters et al., 2014 found that age-related differences in cingulum FA mediated differences in executive functioning. Moreover, white matter changes in the forceps major have been linked to higher performance on working memory tasks (Krogsrud et al., 2018). The remaining tracts (superior longitudinal fasciculus and anterior thalamic radiations) have also been positively correlated with verbal and non-verbal cognitive performance in childhood and adolescence (Tamnes et al., 2010; Urger et al., 2015). I also observed a more surprising negative pathway, between *gc* and the forceps minor in the CALM sample. However, closer inspection showed that the simple association between forceps minor and *gc* was positive, suggesting the negative pathway is likely the consequence of the simultaneous inclusion of collinear predictors (see Tu et al., 2008).

Finally, using SEM trees (Brandmaier et al., 2013), I observed that white matter contributions to *gc* and *gf* differed between participants of different ages (Table 3). In CALM, the contributions of the cingulum, superior longitudinal fasciculus, and forceps major weakened with increasing age for *gf*. For *gc*, however, the forceps major and forceps minor, and the anterior thalamic radiations exhibited a more complex pattern with each tract providing significantly different effects on crystallised intelligence at two distinct time points in development. In NKI-Rockland, the superior longitudinal fasciculus became less associated with both *gc* and *gf*. Considering that decreases in white matter relations to *gc* and *gf* occurred before covariance decreases were found between *gc* and *gf* suggest that differences in white matter development may underlie subsequent individual differences in cognition.

Overall, these findings align with a neurocognitive interpretation of age differentiation-dedifferentiation hypothesis, which would predict that cognitive abilities and their neural substrates become more differentiated (less correlated) until the onset of maturity, followed by an increase (dedifferentiation) in relation to each other until late adulthood (Hartung et al., 2018). However, I must note that the evidence for age differentiation/dedifferentiation was not always robust across analyses methods or samples, suggesting only small effect sizes.

Limitations of the present study

First and foremost, all findings here were observed in cross-sectional samples. To better understand effects such as age differentiation and dedifferentiation, future studies will need to model age-related changes within the same individual. The complexity and expense of collecting such longitudinal data has long precluded such investigations, but new cohorts such as the ABCD sample (Volkow et al., 2018) will allow researchers to model longitudinal changes in the future (see Chapter Four for longitudinal analyses in the IMAGEN study). Secondly, since the tasks modelled here were not identical between cohorts, detailed interpretations of similarities and differences between the CALM and NKI-Rockland samples should be treated with caution. Therefore, future research comparing cohorts may want to prioritise cohorts with matching tasks to maximise comparability. Thirdly, although most findings are similar across our cohorts, some differences were observed, particularly in white matter effects. Although the findings in the SEM tree analysis of age-related differences in white matter to cognition mapping are both cross-validated as well as corrected for multiple comparisons, they remain inherently exploratory. Furthermore, while these findings largely generalise across the two cohorts studied here, further work in larger (such as ABCD, Volkow et al., 2018), more age-heterogeneous (e.g., the Developing Human Connectome Project, Makropoulos et al., 2018) is needed to assess the robustness of these findings. The samples here are considerably larger than typical in the field (Poldrack et al., 2017)—however, even larger samples are desirable to gain truly precise estimates of the key parameters, especially regarding measures such as DTI in the NKI-Rockland sample that have a non-trivial proportion of missing data, which are known to inflate effect sizes (Gelman and Carlin, 2014; Vul et al., 2009). Moreover, the white matter differences observed could also be due to the scans being obtained at different scanner sites, although this is unlikely to have produced considerable differences for all raw images were processed using the same pipeline, and previous work suggests that FA is quite a robust measure in multi-site comparison (see Vollmar et al., 2010).

In terms of analytical frameworks, here a relatively new analytical framework, called SEM trees (Brandmaier et al., 2013), was implemented to allow for recursive partitioning of the cohorts into age-demarcated subgroups to capture developmental heterogeneity. SEM trees have a number of strengths, including considerable flexibility in model specification,

implementation in open-source software, and the ability to combine measurement and structural model components as well as multiple simultaneous predictors. However, they also have challenges, including potential vulnerability to small fluctuations and overfitting (which may cascade down affecting other partitions), and are certainly not the only choice available to examine model heterogeneity. Alternative analytical strategies, varying in the degree to which they presuppose known group membership or estimate it, include finite mixture models (Zadelaar et al., 2019), Gaussian process structural equation models (Silva and Gramacy, 2010), latent class and latent profile analysis (Oberski, 2016), general frameworks such as decision trees (McArdle, 2013) and model-based cluster analysis (Fraley and Raftery, 1999), as well as extensions of SEM trees such as SEM forests (Brandmaier et al., 2016). All these techniques differ in their strengths and weaknesses, ease of implementation, degree of confirmation versus exploration and their flexibility (e.g., can they accommodate latent variables or not). One particularly fruitful avenue for future research is to combine both, using exploratory as well as confirmatory methods to balance discovery and robustness. Here I hopefully illustrate how SEM trees can be one such tool, but would urge the reader to tailor their analytical framework to the question at hand, and be mindful of potential drawbacks. Nonetheless, my view is that SEM trees offer at least one fruitful avenue to formalise hypotheses in developmental cognitive neuroscience which would otherwise often remain mostly verbal.

Related to this point, the findings are further limited by the selection of cognitive tasks for both cohorts. Although the battery of tests used in the current study span three cognitive domains (i.e., crystallised and fluid intelligence, and working memory), inclusion of additional tasks measuring abilities such as processing speed and other executive functions would capture a fuller picture of individual differences in intelligence. For example, including tests of processing speed in the CALM and NKI-Rockland samples would increase the number of indicators (and latent constructs) that can be related to *gc* and/or *gf*. Furthermore, while Matrix Reasoning is the only fluid task identical between both cohorts, NKI-Rockland also contains Block Design, Similarities, and Verbal Reasoning (see Fuhrmann et al., 2020), which were not considered and, therefore, not loaded onto *gf* in the SEM or SEM trees models in this study. Finally, in the CALM sample, the working memory latent variable could be partitioned into traditional working memory (forward and backward recall) and short-term memory (Dot

Matrix and Mr X) to align with more canonical psychometric memory assessments. Taken together, this proposed expanded battery of intelligence-related tasks coupled with more standardised assignment of cognitive tasks to latent constructs would further improve inference for the SEM and SEM trees models as well as provide a more comprehensive and robust estimation of age (de)differentiation between *gc* and *gf* and their associations with white matter (i.e., FA).

Lastly, CALM consists of children with referrals for any difficulties related to learning, attention or memory (Holmes et al., 2019). It should be noted that, since CALM is a sample of children and adolescents struggling to learn, and, therefore, 'atypical', a large percentage of this cohort had been assigned a diagnosis (36.12%). However, controlling for this possible confound through constrained multigroup models showed this did not affect the results of the models, as was seen in previous work using CALM (Fuhrmann et al., 2020). The NKI-Rockland sample, in contrast, is a United States population representative sample (Nooner et al., 2012). Both samples are composed of large cohorts that underwent extensive phenotyping and population-specific representative sampling. Therefore, I argue that the results generalise to 'typical' and 'atypical' samples of neurocognitive development, although further research is required to substantiate this claim.

Summary of the Chapter

In this chapter, I presented analyses that suggested that crystallised and fluid intelligence factors explained a significant amount of variance in test performance in two large child and adolescent samples. These results were found in both typically and atypically (struggling learners) developing cohorts, demonstrating the generalised notion that cognitive ability is *better understood* as a two-factor (*gc* and *gf*) rather than a single-factor (*g*) phenomenon in childhood and adolescence. The addition of white matter microstructure indicated *independent contributions* from specific white matter tracts known to be involved in cognitive ability. Moreover, further analyses suggested that the associations between neural and behavioural measures *differed* during development.

Overall, these results support a neurocognitive age differentiation-dedifferentiation hypothesis (reorganisation) of cognitive abilities whereby the relation between white matter and cognition become *more differentiated* (less correlated) in pre-puberty and then *dedifferentiate* (become more correlated) during early puberty. However, structural equation modelling, although a highly informative and flexible method, assumes the presence of (unobserved) latent variables that cause the variation among cognitive scores. Therefore, in Chapter Three, I now describe analyses done, again using the CALM sample, but this time using the tools of network science, which estimates *interactions* among *observed* variables rather than presuming the existence of unobserved entities. Here I use this approach to model intelligence and its structural brain correlates (i.e., grey and white matter) as a complex system—coinciding with the mutualism theory of cognitive ability (see Chapter One).

Chapter Three

Bridging Brain and Cognition: A Multilayer Network Analysis of Brain Structural Covariance and General Intelligence in the CALM Sample

Chapter Three is available as a preprint (although appears in this thesis in a modified form to ensure consistent formatting and thesis coherence):

Simpson-Kent I. L., Fried E.I., Akarca D., Mareva S., Bullmore E.T., the CALM Team, Kievit R. A. (2021). Bridging brain and cognition: A multilayer network analysis of brain structural covariance and general intelligence in a developmental sample of struggling learners. Accepted, *Journal of Intelligence*. Preprint: *Biorxiv*. doi: 10.1101/2020.11.15.383869

Although first-person pronouns are used throughout the chapter, this work is the result of a collaborative research project. My contributions to the above publication are:

1. I led conceptualisation and planning (aided by R. A. Kievit) about the scientific hypotheses, analysis methods, and interpretations of the project.
2. I performed all manuscript analyses (aided by R. A. Kievit, Eiko I. Fried, and S. Mareva).
3. I wrote the first full draft (with input from R. A. Kievit) of the manuscript and led the revisions (aided by other co-authors) of the manuscript for upload as a preprint in *Biorxiv*.

Introduction

In Chapter Two, I took a traditional approach to studying associations between intelligence-related measures: structural equation modelling (SEM). Although the use of SEM trees is novel and more data-driven than more widely used SEM methods, it uses a confirmatory factor modelling approach, which posits the existence of a latent (unobserved) variable that causes variation in the observed scores. However, SEM is not the only conceptual or methodological approach to study intelligence and its neural correlates. Other perspectives, for example, instead posit direct connections between cognitive domains. In this chapter, I use network analysis to study the associations between crystallised and fluid intelligence, working memory, and brain structure (cortical volume and fractional anisotropy) in the CALM sample (ages 5 – 18 years). Rather than opposing the results of Chapter Two, this network model of intelligence (i.e., mutualism) complements them and previous research. However, network science conceptualises intelligence as a *complex system* and uses exploratory methods (i.e., partial correlations and community detection), *without* presuming the influence of latent variables.

In recent years, methods from network analysis have shed new light on both the cognitive abilities that make up general intelligence (Kievit et al., 2019; van der Maas et al., 2017), as well as the brain systems purported to support them (Girn et al., 2019; Seidlitz et al., 2018). For instance, the mutualism model (van der Maas et al., 2006) was inspired by an ecosystem model of prey-predator relations, and states that the positive manifold (Spearman, 1904), rather than existing in final form since birth, *emerges* gradually from the *positive interactions* among different cognitive abilities (i.e., reasoning and vocabulary) over time (see Kievit et al., 2019, 2017). Hence, the positive manifold (and hence, general intelligence) can arise even from originally weakly correlated cognitive faculties. The mutualism model (also see van der Maas et al., 2017) therefore highlights the need to both conceptualise traits, abilities, or psychological constructs such as general intelligence as complex dynamical systems, as well as use appropriate statistical models (i.e., network analysis) to estimate relationships among elements of the systems under investigation (Fried, 2020; Fried and Robinaugh, 2020).

For instance, new innovations in network psychometrics (Epskamp et al., 2018) have led to a rapid increase in popularity of behavioural network analysis, especially in

psychopathology (Borsboom, 2017; Robinaugh et al., 2019). In this framework, psychological constructs are theorised as complex systems, whereby relationships (edges) between nodes (e.g., item responses on a questionnaire) are estimated using weighted *partial correlation networks*. This use of partial correlations enables the determination of *conditional dependencies* among variables, *after* controlling for the associations among every other node in the network (Epskamp et al., 2018).

This approach has also recently been used to analyse cross-sectional data on general intelligence. For instance, both Kan et al., 2019 (N = 1,800; age range: 16 – 89 years) as well as Schmank et al., 2019 (N = 1,112; age range: 12 – 90 years) used a network model approach to analyse data from the WAIS-IV cognitive battery (Wechsler, 2008). The network model showed better fit to the pattern of intelligence scores compared to a latent variable approach (*g* factor), in support of mutualism. Furthermore, Mareva and Holmes, 2020, in two separate samples, one the same group of struggling learners as studied here (CALM) but with fewer participants (N = 350), no neuroimaging data, and including tasks not analysed in this study (e.g., motor speed and tower achievement), observed links between cognitive abilities and learning, especially between mathematics skills and more 'domain-general' faculties such as backward digit span and matrix reasoning.

Besides psychology, in neuroscience, network analysis methods have been widely used to describe the relations among brain regions, ushering in the field of network neuroscience (Bassett and Sporns, 2017; Fornito et al., 2016). Rather than focusing on individual brain regions in isolation, the brain is conceived as a complex system of interconnected networks that facilitate behavioural functions ranging from sensorimotor control to learning. Several influential studies have revealed pervasive properties of brain networks that enable adaptive behaviour such as small-world topology (Bassett and Bullmore, 2006, 2017), modularity (Meunier et al., 2010; Sporns and Betzel, 2016) and 'rich-club' connector hubs (Heuvel and Sporns, 2011), consistent with an economical trade-off between minimising wiring cost and maximising efficiency (e.g., information transfer) (Bullmore and Sporns, 2012). Furthermore, in the same sample studied here, Akarca et al., 2020 applied a generative network modelling approach to simulate the growth of brain network connectomes. The findings demonstrate the possibility of simulating structural networks with statistical properties mirroring the spatially embedding of those observed in the real-life

brains. The parameters of these generative models were shown to correlate with neuroimaging measures not used to train the models (including grey matter measures), cognitive performance (including vocabulary and mathematics) and relate to gene expression in the cortex. Together these studies point the field toward a better mechanistic understanding of the development of human brain structure, function, and their relationship with cognitive ability.

Despite the success of network approaches in providing unique insights within both cognitive psychology and cognitive neuroscience, few studies have integrated them into a so-called *multilayer network* paradigm (Bianconi, 2018), which models the relationships among variables *simultaneously* across time (e.g., days, weeks, months, and years) and/or levels of organisation (e.g., behaviour and brain variables). Two studies have recently pushed this boundary. Hilland et al., 2020 examined the relations between brain structure (cortical thickness and volume) and depression symptoms. They found (via a partial correlation network model) that certain clusters of brain regions (i.e., the cingulate, fusiform gyrus, hippocampus, and insula) were conditionally dependent with a subset of depression symptoms (crying, irritability and sadness). Secondly, in 172 male autistic participants (ages 10 – 21 years), Bathelt et al., 2020 used 'network-based regression' to estimate the relationship between the unique variance of both the autism symptom network and functional brain connectivity (resting-state fMRI). Moreover, they applied Bayesian network analysis to create a directed acyclic graph between subscores of autism symptoms and their neural correlates. They found that communication and social behaviour were predicted by their respective resting-state MRI neural correlates (termed 'Comm Brain' and 'Social Brain').

In Chapter Two, I showed in two large samples (CALM and NKI-Rockland) that the variance of general intelligence (Spearman, 1904) and its white matter correlates (fractional anisotropy, see Wandell, 2016) were better captured by separate factors (i.e., *gc*, *gf*, and 5 of 10 white matter tracts) rather than a single-factor model of cognition or global white matter (*g*). While factor models have traditionally been used to study intelligence (e.g., Carroll, 1993), in the last two decades there has been a rise in use of the statistical tools of network science (Barabási, 2016) to show that examining relationships between cognitive abilities can help us better understand the development of general intelligence. In this chapter, I use techniques from network science (i.e., graphical LASSO, strength centrality, and the Walktrap

community detection algorithm) to extend previous work assessing the mutualism model of intelligence (Figure 4). I also expand on work in the previous chapter by including grey matter (i.e., cortical volume). I conduct these analyses in the CALM sample (behavioural, N = 805; cortical volume, N = 246; fractional anisotropy, N = 165, age range: 5 – 18 years).

Methods

Participants: The CALM cohort

The present cross-sectional sample (behavioural, N = 805; cortical volume, N = 246; fractional anisotropy, N = 165; age range: 5 to 18 years) was obtained from the Centre for Attention, Learning and Memory (CALM) located in Cambridge, UK (Holmes et al., 2019), which has been described in Chapter Two (**Participants: The CALM and NKI-Rockland cohorts** section).

Cognitive assessments: *gc*, *gf*, and working memory

I analysed the same tasks (measures of crystallised and fluid intelligence, and working memory) outlined in Chapter Two (Table 1). Furthermore, I included the task Following Instructions (mean: 11.2, standard deviation: 4, range: [1, 33], missingness: 6.83%), which was not part of the previous study in Chapter Two. This was done to balance the number of variables analysed for cognitive and brain structure measures in the multilayer networks (i.e., cognition = grey matter = white matter = 10 nodes each). Moreover, doing so helps to prevent biased findings by ensuring that the partial correlation estimation procedures (see below) are applied across an equal number of nodes for each level of the multilayer networks. Following Instructions, a working memory task (Gathercole et al., 2008), requires participants to carry out various sequences of actions (touch and/or pick up) involving objects (a box, an eraser, a folder, a pencil or a ruler), which are presented in front of them. Participants undertake actions sequentially (do X "then" do Y), with increasingly longer instruction sequences, leading to increased difficulty. Performance scores denote total number of correct responses. To view age trends of the performance scores for the cognitive tasks, see Figure 11. It must be noted that, since the present sample size is larger (i.e., 805 vs 551), the descriptive statistics and missingness (range: 0.12% to 9.94%) are different for this study.

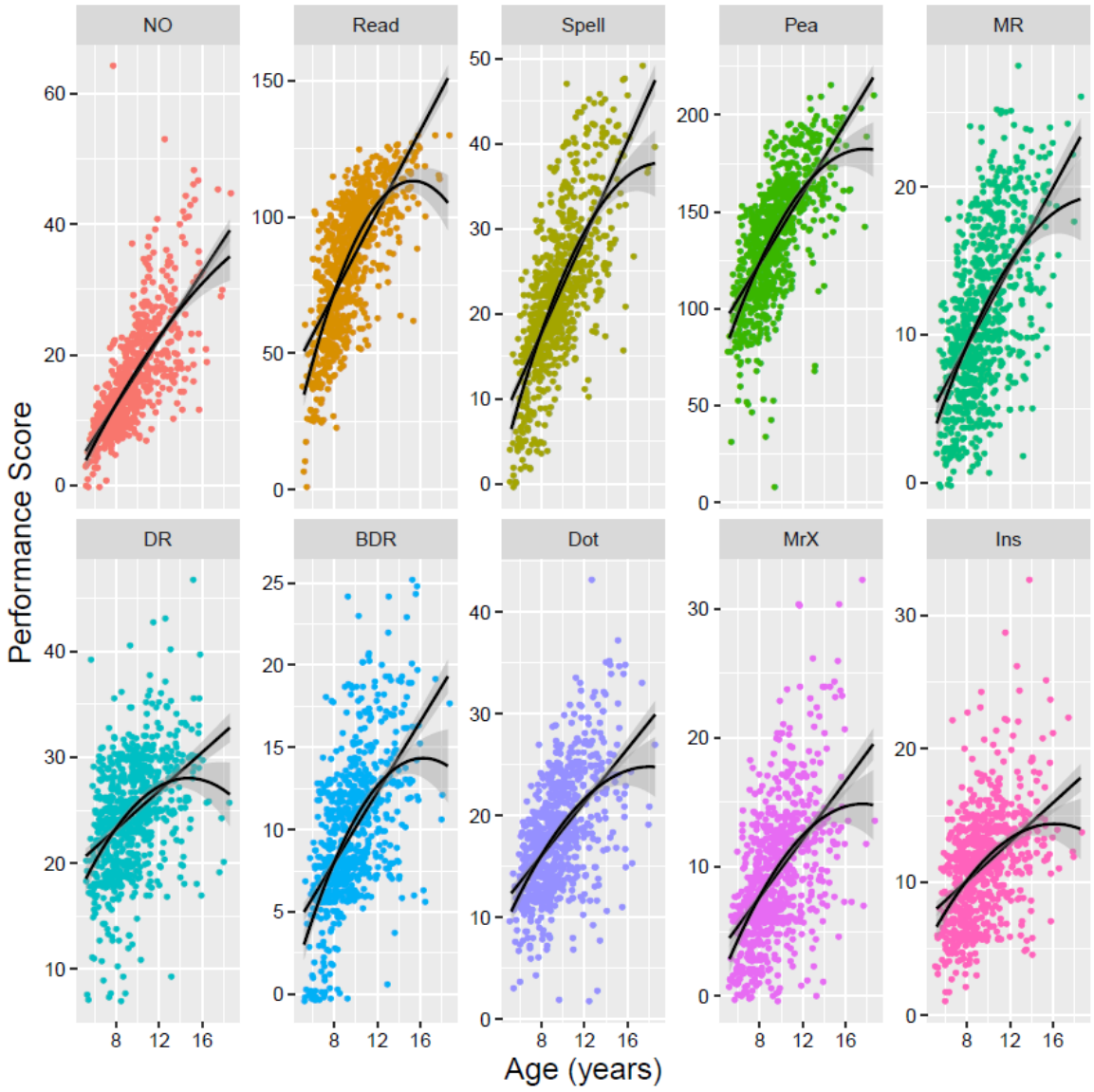


Figure 11. Cross-sectional scatterplot for cognitive raw scores. Solid lines represent linear and polynomial fit while shades indicate 95% confidence intervals. Abbreviations: Matrix Reasoning (MR), Peabody Picture Vocabulary Test (Pea), Spelling (Spell), Single Word Reading (Read), Numerical Operations (NO), Digit Recall (DR), Backward Digit Recall (BDR), Mr. X (MrX), Dot Matrix (Dot), and Following Instructions (Ins). Note, that this figure differs from Figure 8 by including a larger sample size (805 vs 551) and the Following Instructions task.

Structural MRI measures: Cortical volume (CV) and fractional anisotropy (FA)

All CALM T1-weighted volume scans and Diffusion-Weighted Images (DWI) acquisition protocols, parameters, MRI quality control procedures (to minimise potential biases due to subject movement), and steps/pipelines to compute regional CV estimation and

FA maps were the same as described in Chapter Two (section **Structural MRI measures: Fractional anisotropy (FA)**).

As the grey matter metric, I used region-based cortical volume (CV, in mm^3; N = 246, averaged across contralateral homologues), based on the Desikan-Killiany atlas (Desikan et al., 2006) and defined as the distance between the outer edge of cortical grey matter and subcortical white matter (Fischl and Dale, 2000). Tissue classification and anatomical labelling was performed based on the T1-weighted scan using FreeSurfer v5.3.0 software (see http://surfer.nmr.mgh.harvard.edu/ for free download and documentation) The technical details of these procedures are described in prior publications (Dale et al., 1999; Fischl et al., 2002, 1999). FreeSurfer morphology output statistics were computed for each ROI, and also included cortical thickness and surface area (see **Supplementary Material for Chapter Three**, section **Teasing apart the relations of cortical volume to general intelligence: Multilayer analysis using cortical surface area and thickness** for analyses involving these two metrics). I included a subset of 10 cortical volume regions in this study: caudal anterior cingulate (CAC), caudal middle frontal gyrus (CMF), frontal pole (FP), medial orbitofrontal cortex (MOF), rostral anterior cingulate gyrus (RAC), rostral middle frontal gyrus (RMF), superior frontal gyrus (SFG), superior temporal gyrus (STG), supramarginal gyrus (SMG), and transverse temporal gyrus (TTG). Moreover, for fractional anisotropy (FA, N = 165), a proxy measure for white matter integrity (Wandell, 2016), I included 10 regions using the Johns Hopkins University DTI-based white matter tractography atlas (see Hua et al. 2008): anterior thalamic radiations (ATR), corticospinal tract (CST), cingulate gyrus (CING), cingulum [hippocampus] (CINGh), forceps major (FMaj), forceps minor (FMin), inferior fronto-occipital fasciculus (IFOF), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), and uncinate fasciculus (UNC).

I used these region-based measures to study brain structural covariance (Alexander-Bloch et al., 2013), which have been used in cross-sectional and longitudinal designs of cognitive ability in childhood and adolescence (e.g., Solé-Casals et al., 2019; see Kievit and Simpson-Kent, 2021 for a recent review of longitudinal studies). In addition, emerging theoretical proposals emphasise the role of networks of brain areas in producing intelligent behaviour (e.g., Parieto-Frontal Integration Theory (P-FIT), Jung and Haier, 2007, and The Network Neuroscience Theory of Human Intelligence, Barbey, 2018) rather than individual

regions-of-interest (ROIs) in isolation (e.g., primarily the prefrontal cortex). I selected these 10 grey matter and 10 white matter ROIs based upon combined evidence from a recent meta-analysis (Basten et al., 2015) on associations between functional and structural ROIs and cognitive ability that further extended the P-FIT theory, but also more recent work involving two large cohorts, one in longitudinal analysis of the UK Biobank sample (grey matter, Kievit et al., 2018b) and the other using the same (cross-sectional) CALM sample in Chapter Two (cognitive data, N = 551; white matter data, N = 165).

To view age trends of CV and illustrations of ROIs, see Figure 12. For ROI visualisations and ages trends of FA, see Figure 9 of Chapter Two. For correlation plots of cognitive tasks and neuroimaging measures, see Figure 13.

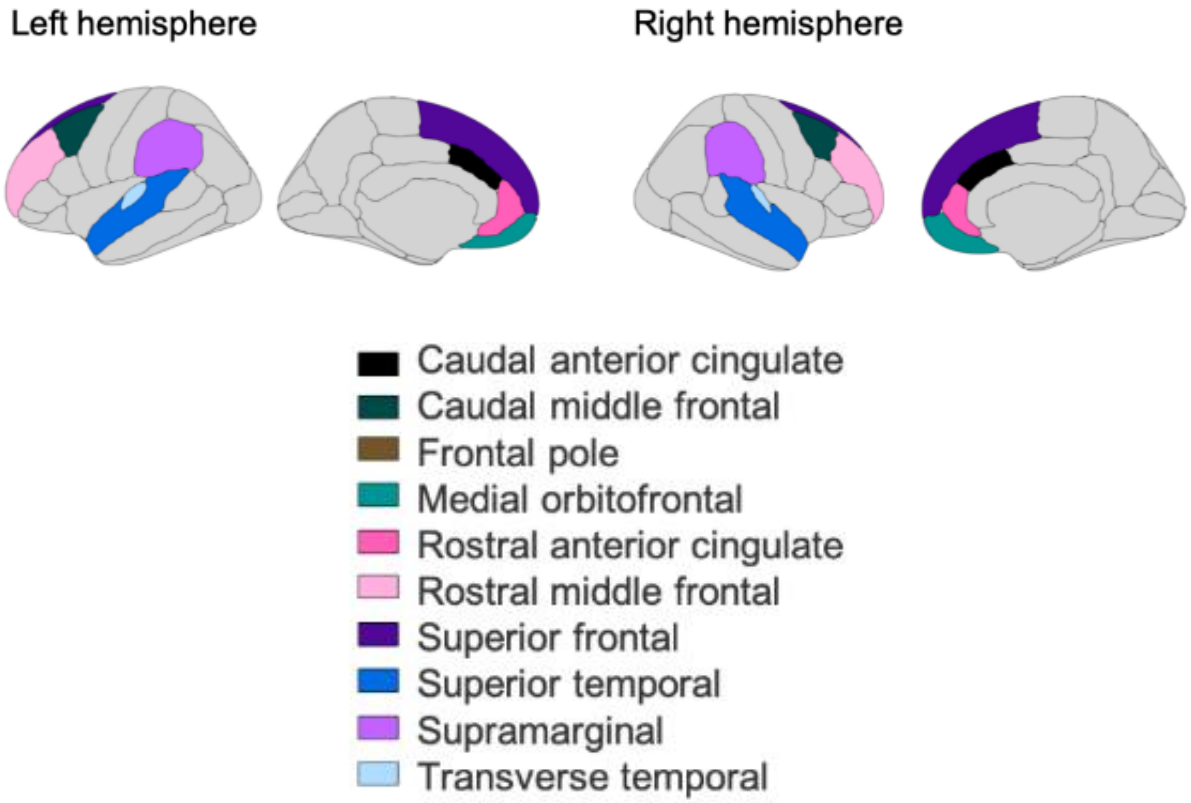


Figure 12. Top: Grey matter ROIs based on the DK atlas (cortical volume, N = 246) in the left and right hemisphere.

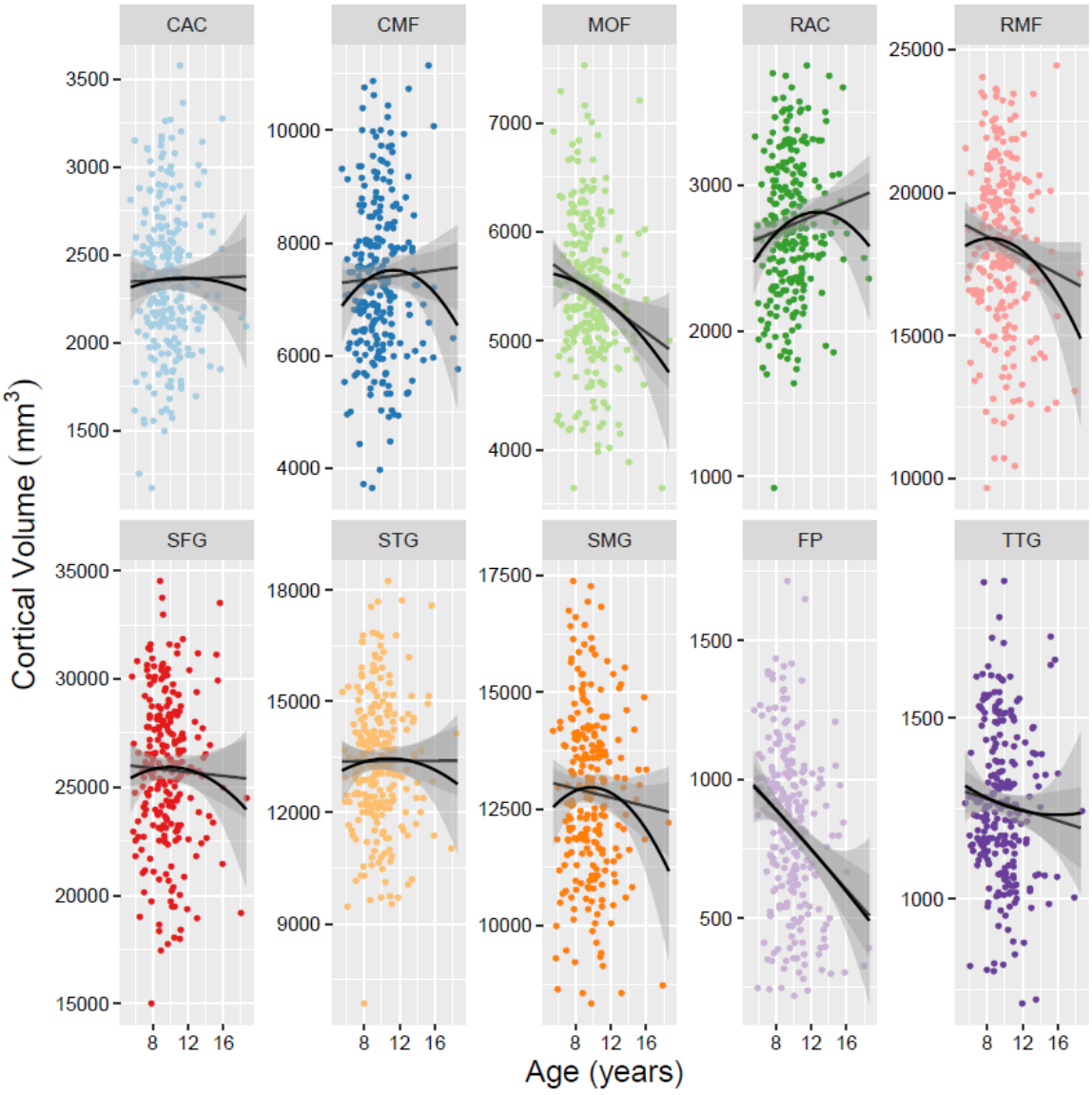


Figure 12. Bottom: Cross-sectional scatterplot for bilateral cortical volume. Solid lines represent linear and polynomial fit while shades indicate 95% confidence intervals. Abbreviations: caudal anterior cingulate (CAC), caudal middle frontal gyrus (CMF), medial orbitofrontal cortex (MOF), rostral anterior cingulate gyrus (RAC), rostral middle frontal gyrus (RMF), superior frontal gyrus (SFG), superior temporal gyrus (STG), supramarginal gyrus (SMG), frontal pole (FP), and transverse temporal gyrus (TTG).

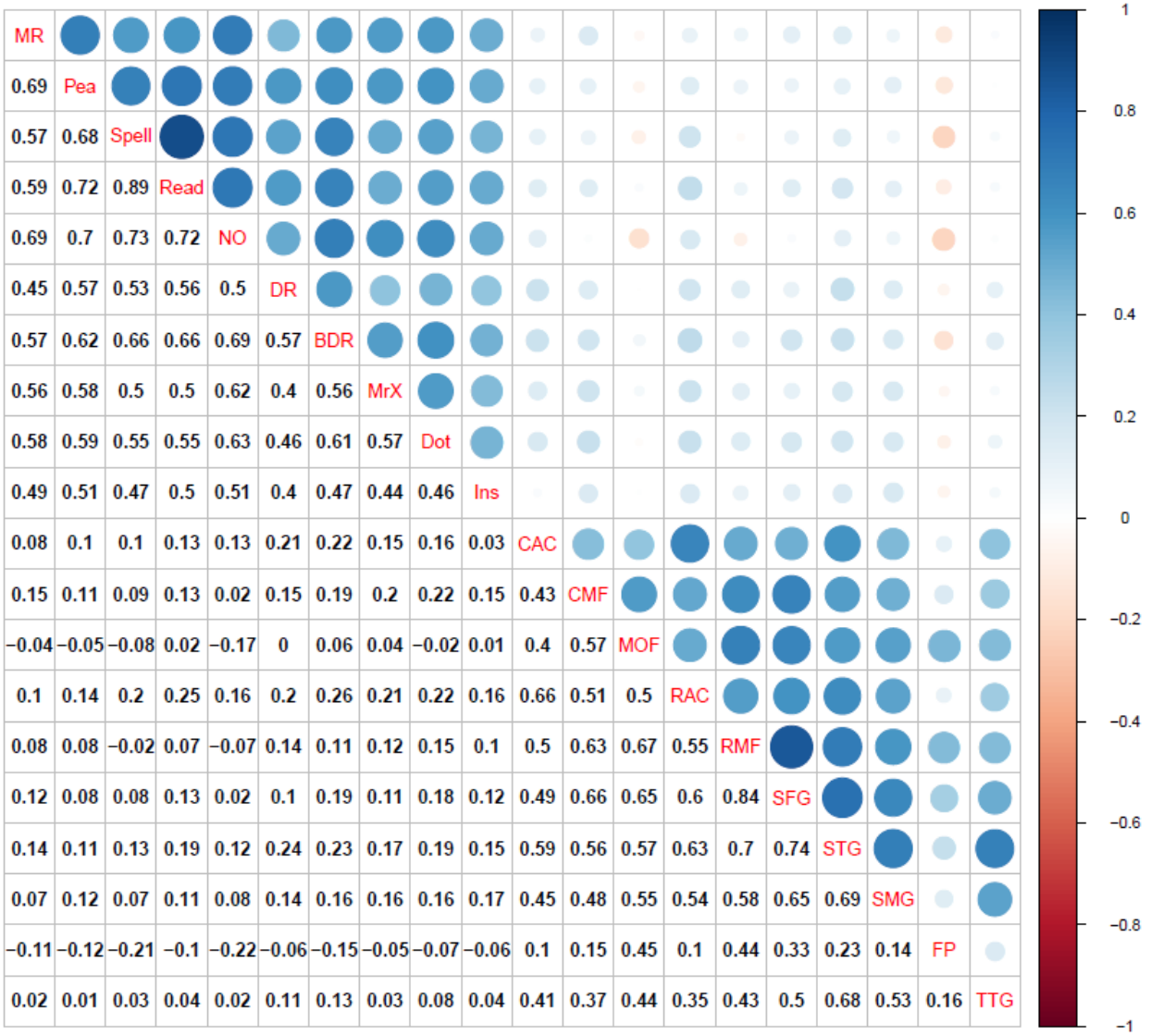


Figure 13 Top. Correlation plot for cognitive raw scores and bilateral cortical volume ROIs. All coefficients shown are Pearson correlations. Blue represents positive correlations while red signifies negative correlations among variables. Size of circles indicates the magnitude of the association (e.g., larger circle = higher correlation). Correlations calculated using pairwise complete observations. Abbreviations: Matrix Reasoning (MR), Peabody Picture Vocabulary Test (Pea), Spelling (Spell), Single Word Reading (Read), Numerical Operations (NO), Digit Recall (DR), Backward Digit Recall (BDR), Mr. X (MrX), Dot Matrix (Dot), Following Instructions (Ins), caudal anterior cingulate (CAC), caudal middle frontal gyrus (CMF), medial orbitofrontal cortex (MOF), rostral anterior cingulate gyrus (RAC), rostral middle frontal gyrus (RMF), superior frontal gyrus (SFG), superior temporal gyrus (STG), supramarginal gyrus (SMG), frontal pole (FP), and transverse temporal gyrus (TTG).

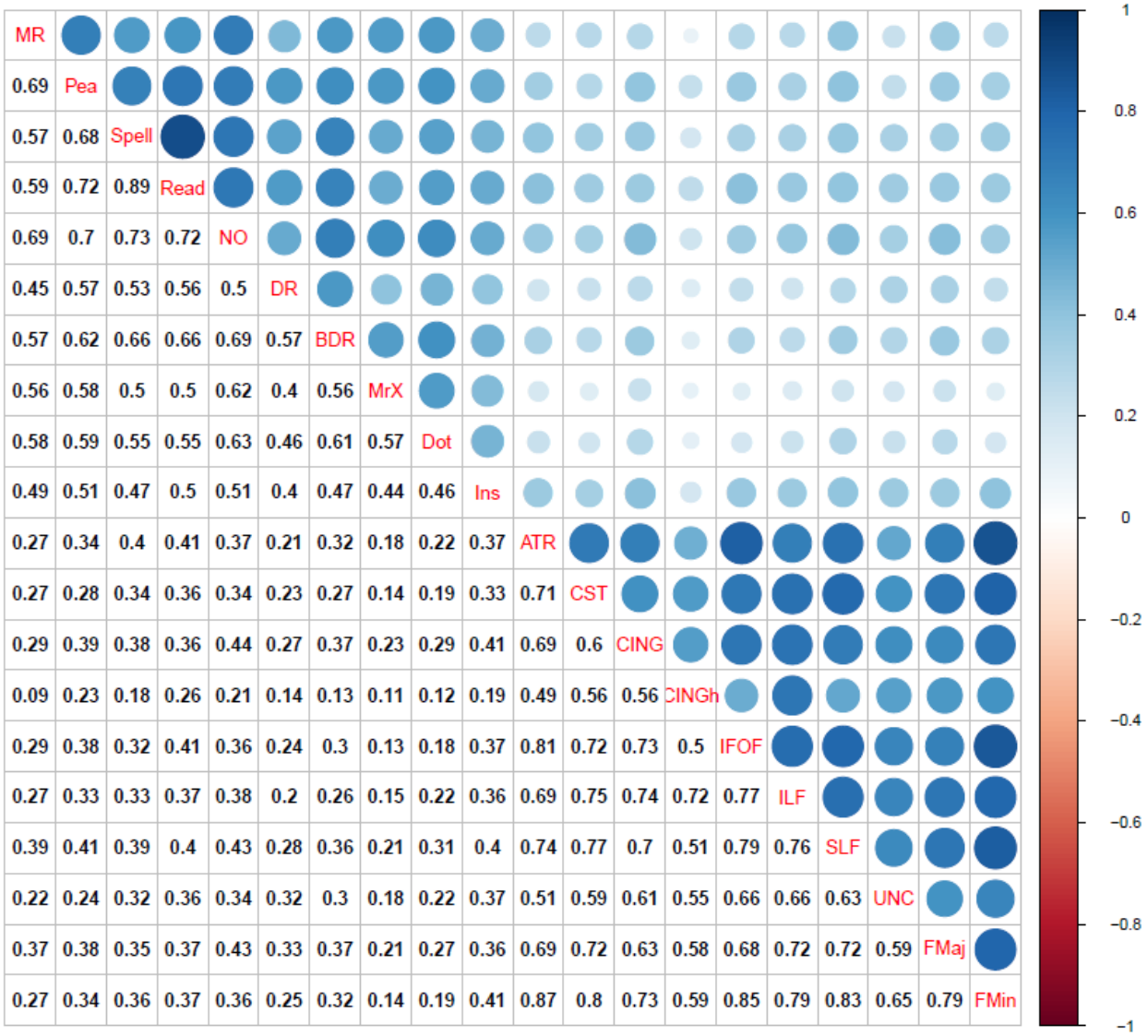


Figure 13 Bottom. Correlation plot for cognitive raw scores and bilateral fractional anisotropy ROIs. All coefficients shown are Pearson correlations. Blue represents positive correlations while red signifies negative correlations among variables. Size of circles indicates the magnitude of the association (e.g., larger circle = higher correlation). Correlations calculated using pairwise complete observations. Abbreviations: Matrix Reasoning (MR), Peabody Picture Vocabulary Test (Pea), Spelling (Spell), Single Word Reading (Read), Numerical Operations (NO), Digit Recall (DR), Backward Digit Recall (BDR), Mr. X (MrX), Dot Matrix (Dot), Following Instructions (Ins), anterior thalamic radiations (ATR), corticospinal tract (CST), cingulate gyrus (CING), cingulum [hippocampus] (CINGh), inferior fronto-occipital fasciculus (IFOF), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), uncinate fasciculus (UNC), forceps major (FMaj), and forceps minor (FMin).

Statistical analyses: Network analyses and community detection

All statistical analyses and plots were performed using R (R Core Team, 2020) version 3.6.3 ("Holding the Windsock"). Network estimation was performed using the packages bootnet (version 1.4.3, Epskamp and Fried, 2020) , igraph (version 1.2.6, Amestoy et al., 2020), qgraph (version 1.6.5, Epskamp et al., 2020), and networktools (version 1.2.3, Jones, 2020). I used these tools to estimate weighted partial correlation networks, which allowed determination of conditional dependencies among the cognitive and neural variables. For example, in a multilayer network, any partial correlation between node A (e.g., Matrix Reasoning) and node B (e.g., the caudal anterior cingulate) is one that remains after controlling for the associations among A and B with every other node in the network (e.g., other cognitive abilities and cortical volume ROIs). To estimate these networks, I applied Gaussian Graphical Models (Pearson correlations) using regularisation (graphical lasso, see Friedman et al., 2008) with a threshold tuning parameter of 0.5 and pairwise deletion to account for missingness. These methods have been widely used to generate sparser networks by penalising for more complex models—thus, decreasing the risk of potentially spurious (e.g., false positive) connections and enabling simpler visualisation and interpretation of conditional dependencies between nodes (Epskamp and Fried, 2018). I hypothesised that the results would show positive partial correlations (in line with mutualism theory) both within cognitive (e.g., as observed in Mareva and Holmes, 2020 and Schmank et al., 2019) and within neural measures (single-layer networks) as well as between brain-behaviour variables in the multilayer networks.

Age was included as a node in the estimation procedures of all partial correlation networks (i.e., edge weights, centrality, network stability, and community detection) but was not included in the visualisations of the networks and centrality plots, or in network descriptive statistics (i.e., mean, median, and range of edge weights). For a comparison and discussion of the use of age (i.e., included in network estimation, regressed out beforehand, or removed from dataset before network estimation), see the **Supplementary Material for Chapter Three** section **How to deal with age?**.

To assess the statistical interconnectedness or connectivity of cognitive and neural nodes relative to their neighbours within the single-layer networks, I estimated node strength, a weighted degree centrality measure calculated by summing the absolute partial correlation

coefficients (edge weights) between a node and all other nodes it connects to within the network. Note that the brain structural covariance networks involve ROIs that are not necessarily anatomically connected, preventing certain inferences such as information flow. Nodes were classified as central if the magnitude of their strength z-score was positive and equal to or greater than one standard deviation above the mean. I do not discuss or interpret negative centrality z-score values for the single-layer networks.

In the multilayer networks, I applied the Walktrap community detection algorithm (Pons and Latapy, 2005) to determine in a data-driven manner whether clustering, or grouping, of nodes (e.g., cognitive and/or neural) occurred. The Walktrap algorithm assesses how strongly related nodes are to each other (that can be due to similarity, e.g., because nodes A and B are similar, or it can be because nodes A and B are different but node A has a strong impact on node B; see "*Topological overlap and missing nodes*" of Fried and Cramer, 2017). The Walktrap algorithm takes recursive random walks between node pairs and classifies communities according to how densely connected these parts are within the network (wherever the random walks become 'trapped'). Walktrap is widely used in the network psychometrics literature and, in a Monte Carlo simulation study, was shown to outperform other algorithms (e.g., InfoMap) for sparse count networks (e.g., those used in diffusion tensor imaging). However, it must be noted that this result was found for networks made up of 500 nodes or higher (Gates et al., 2016). I also calculated the maximum modularity index value (Q), which estimates the robustness of the community partition (Newman, 2006). I interpreted values of 0.5 or above as evidence for reliable grouping.

Instead of traditional absolute strength, for the multilayer networks I calculated bridge strength, a novel weighted degree centrality measure originally developed to study comorbidity between mental disorders (see Jones et al., 2019 for overview). Bridge strength centrality sums the absolute value of every edge that connects one node (e.g., Matrix Reasoning) in one pre-assigned community (e.g., cognition) to another node (e.g., caudal anterior cingulate) in another pre-assigned community (e.g., brain). Recent simulation work has shown that the method can reliably recover true structures of bridge nodes in both directed and undirected networks (Jones et al., 2019). Rather than relying on straightforward 'brain' or 'behaviour' assignments to classify nodes, I pre-assigned communities for bridge strength calculation based on results from the Walktrap algorithm.

The presence of bridges between communities (e.g., if nodes from topological distinct clusters such as cognition vs brain feature relations) might suggests the existence of intermediate endophenotypes (Fornito and Bullmore, 2012; Kievit et al., 2016), and potentially identify nodes (both cognitive and neural) that might one day guide intervention studies. Nodes were classified as central if the magnitude of their bridge strength z-score was positive and equal to or greater than one standard deviation above the mean. I do not discuss or interpret negative centrality z-score values for the multilayer networks.

Finally, I quantified the reliability of the centrality estimates for all single-layer (absolute strength of cognitive and brain structural covariance nodes) and multilayer networks (bridge strength). To do this, I estimated the correlation stability (CS)-coefficient. The CS-coefficient calculates the maximum proportion (out of 2000 bootstraps) of the sample that can be dropped out and, with 95% probability, still retain a correlation of 0.7 (correlation between rank order of centrality in network estimated on full sample with order of subsampled network in smaller N) (Epskamp et al., 2018). A CS-coefficient value of 0.5 is considered to be stable. Lastly, also using bootstrapping, I determined the stability of the edge-weight coefficients but present these results in the **Supplementary Material for Chapter Three** of this dissertation (section **Edge-weight stability analyses**).

Results

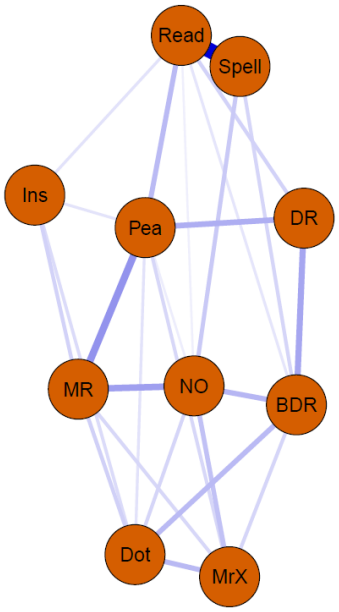
Single-layer network models (cognitive, cortical volume, and fractional anisotropy)

The regularised partial correlation (PC) network for the CALM cognitive data is shown in Figure 14 (top left). This network shows that all partial correlations are positive, and most have small magnitude (mean PC = 0.08, median PC = 0.07, PC range = 0 – 0.63). One edge (between Reading and Spelling) was an outlier (PC = 0.63, all others are between 0 and 0.27), likely due to close content overlap (verbal ability). Regarding centrality, three nodes emerged as strong (positive z-score at or greater than one standard deviation above the mean): (in descending order of centrality strength) Reading, Numerical Operations, and Peabody Picture Vocabulary Test (Figure 14 top right). Overall, centrality estimates were stable, indicated by a high correlation stability (CS)-coefficient of 0.75, revealing that at least 75% of the sample could be dropped while maintaining a correlation of 0.7 with the original sample at 95% probability.

Next, I estimated the partial correlation network among the 10 grey matter regions as shown in Figure 12 above. All edge weights (mean PC = 0.09, median PC = 0, PC range = -0.15 – 0.52) of the cortical volume network (Figure 14, middle left) were positive apart from one negative path (caudal middle frontal gyrus and frontal pole PC = -0.15). Note, the negative path between the caudal middle frontal gyrus and frontal pole might be due to the weak correlations between the frontal pole with other grey matter nodes and displaying a steeper decrease pattern across age (Figures 12 and 13). Two ROIs emerged as central (in descending order of centrality strength): superior temporal gyrus and rostral middle frontal gyrus (Figure 14, middle right). Similar to the cognitive network, cortical volume centrality was stable (CS-coefficient = 0.52), indicating that about 52% of the sample could be subtracted to maintain a correlation of centrality estimates above 0.7 compared to the full sample. This finding is despite the lower sample size compared to the behavioural data (N = 805 for behaviour vs N = 246 for cortical volume).

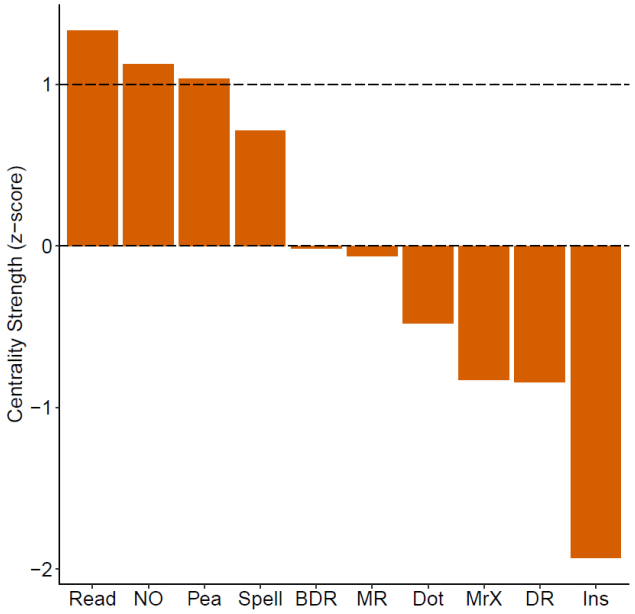
Finally, similar to the cognitive and the grey matter covariance network, the fractional anisotropy network (Figure 14, bottom left) has positive partial correlations with all edge weights varying between small and moderate values: mean PC = 0.08, median PC = 0, and PC range = 0 – 0.44. Two white matter ROIs displayed centrality (Figure 14, bottom right). These included (in descending order) the forceps minor and inferior longitudinal fasciculus. Finally,

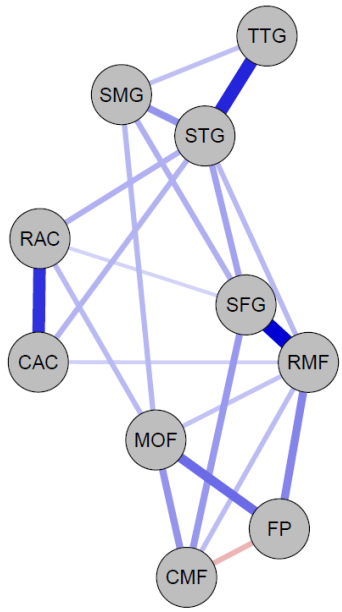
fractional anisotropy centrality was moderately stable (CS-coefficient = 0.44) indicating that about 44% of the sample could be removed while maintaining an association of 0.7 with 95% probability. This is possibly due to the much lower sample size (N = 165) compared to the cognitive (N = 805) and grey matter (N = 246) networks.



Cognition

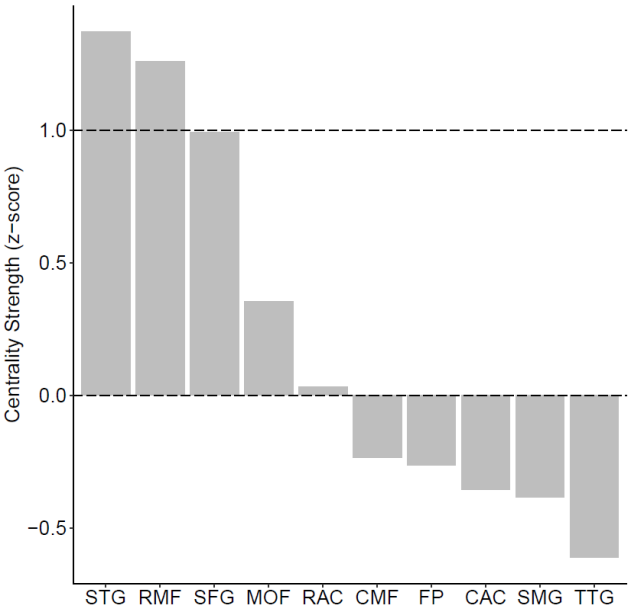
- MR: Matrix Reasoning
- Pea: Peabody Picture Vocabulary Test
- Spell: Spelling
- Read: Reading
- NO: Numerical Operations
- DR: Digit Recall
- BDR: Backward Digit Recall
- MrX: Mr. X
- Dot: Dot Matrix
- Ins: Following Instructions

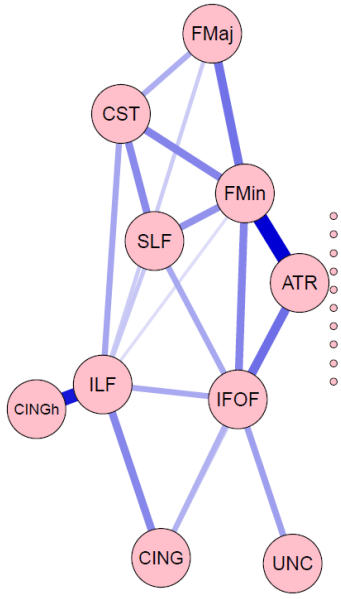




Bilateral Cortical Volume

- CAC: Caudal Anterior Cingulate
- CMF: Caudal Middle Frontal Gyrus
- MOF: Medial Orbitofrontal Cortex
- RAC: Rostral Anterior Cingulate Gyrus
- RMF: Rostral Middle Frontal Gyrus
- SFG: Superior Frontal Gyrus
- STG: Superior Temporal Gyrus
- SMG: Supramarginal Gyrus
- FP: Frontal Pole
- TTG: Transverse Temporal Gyrus





Bilateral Fractional Anisotropy

- ATR: Anterior Thalamic Radiations
- CST: Corticospinal Tract
- CING: Cingulate Gyrus
- CINGh: Cingulum [hippocampus]
- IFOF: Inferior Fronto−Occipital Fasciculus
- ILF: Inferior Longitudinal Fasciculus
- SLF: Superior Longitudinal Fasciculus
- UNC: Uncinate Fasciculus
- FMaj: Forceps Major
- FMin: Forceps Minor

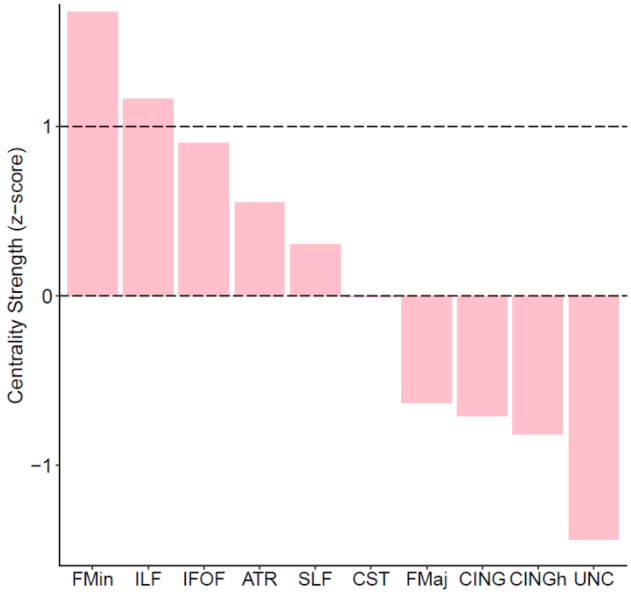


Figure 14. Single-layer partial correlation networks. Top: Network visualisation (spring layout, left side) of CALM cognitive data (N = 805). Centrality estimates (z-scores) of all cognitive tasks (right). Middle: Network visualisation (spring layout, left side) of CALM cortical volume data (N = 246). Centrality estimates (z-scores) of all cortical volume nodes (right). Bottom: Network visualisation (spring layout, left side) of CALM fractional anisotropy data (N = 165). Centrality estimates (z-scores) of all fractional anisotropy nodes (right). Dashed lines indicate mean strength and one standard deviation above the mean.

Bridging the gap: Multilayer networks

The regularised partial correlation network analyses for the CALM multilayer networks data are shown in Figure 15. Consistent with the pattern found in the single-layer networks, the cognitive and grey matter multilayer network (top left of Figure 15) edges are mostly positive with small to moderate weights (mean PC = 0.04, median PC = 0, PC range = -0.12 – 0.64). Comparably, the cognitive and white matter multilayer network (Figure 15, top right) had similar edge weight estimates (mean PC = 0.04, median = 0, range = -0.2 – 0.65). Finally, combining all measures together (tri–layer network consisting of cognition, grey and white matter, bottom centre of Figure 15) produced a network with similar characteristics to the bi-layer networks (mean PC = 0.02, median PC = 0, PC range = -0.2 – 0.66). For the bi-layer networks, the Walktrap algorithm produced either three (cognition-white matter) or four (cognitive-grey matter) clusters that consisted entirely of cognitive or neural nodes except for Following Instructions (Ins), which was either separated from the network (cognition-grey matter, Q = 0.56, indicating strong modularity) or grouped with a neural node (forceps minor of the cognition-white matter network, Q = 0.39, indicating moderate modularity). The result for the tri-layer network (Q = 0.25, indicating weak modularity) was more complex with a total of 15 communities (Figure 15, bottom centre).

Regarding centrality, I report bridge strength (Figure 16). In the cognitive-grey matter network, three bridge nodes surfaced (in descending order: superior temporal gyrus, superior frontal gyrus, and rostral middle frontal gyrus, Figure 16 top left). In terms of stability, the CS-coefficient was 0.20, indicating that the bridge strength estimates were unstable under bootstrapping conditions. In the cognitive-white matter bi-layer network, three nodes (in descending order: uncinate fasciculus, inferior frontal-occipital fasciculus, and hippocampal cingulum) emerged as possible bridge nodes (Figure 16, top right). Moreover, the centrality estimates had a CS-coefficient of 0.13, once again suggesting that the bridge strength estimates

were unstable. Lastly, for the tri-layer network, five nodes displayed positive bridge strength equal to or greater than one standard deviation above the mean (Figure 16, bottom centre). These included (in descending order): Reading, Peabody Picture Vocabulary Test, superior frontal gyrus, Spelling, and Numerical Operations. Much better than the bi-layer networks, the tri-layer network bridge strength estimates were moderately stable (CS-coefficient = 0.44).

Lastly, I re-ran our analyses to test the sensitivity of the main findings (e.g., positive partial correlations and central nodes) to potential outliers (defined as ± 4 standard deviations). Doing so did not severely alter the partial correlation weights between nodes in the networks (see section **The possible effect of outliers on major findings** of **Supplementary Material for Chapter Three** for detailed comparisons). It must be restated that this study's data comes from an atypical sample, which might influence brain metrics even with rigorous quality control procedures. Therefore, despite this discrepancy, the data supports brain-behaviour 'bridges' in general intelligence.

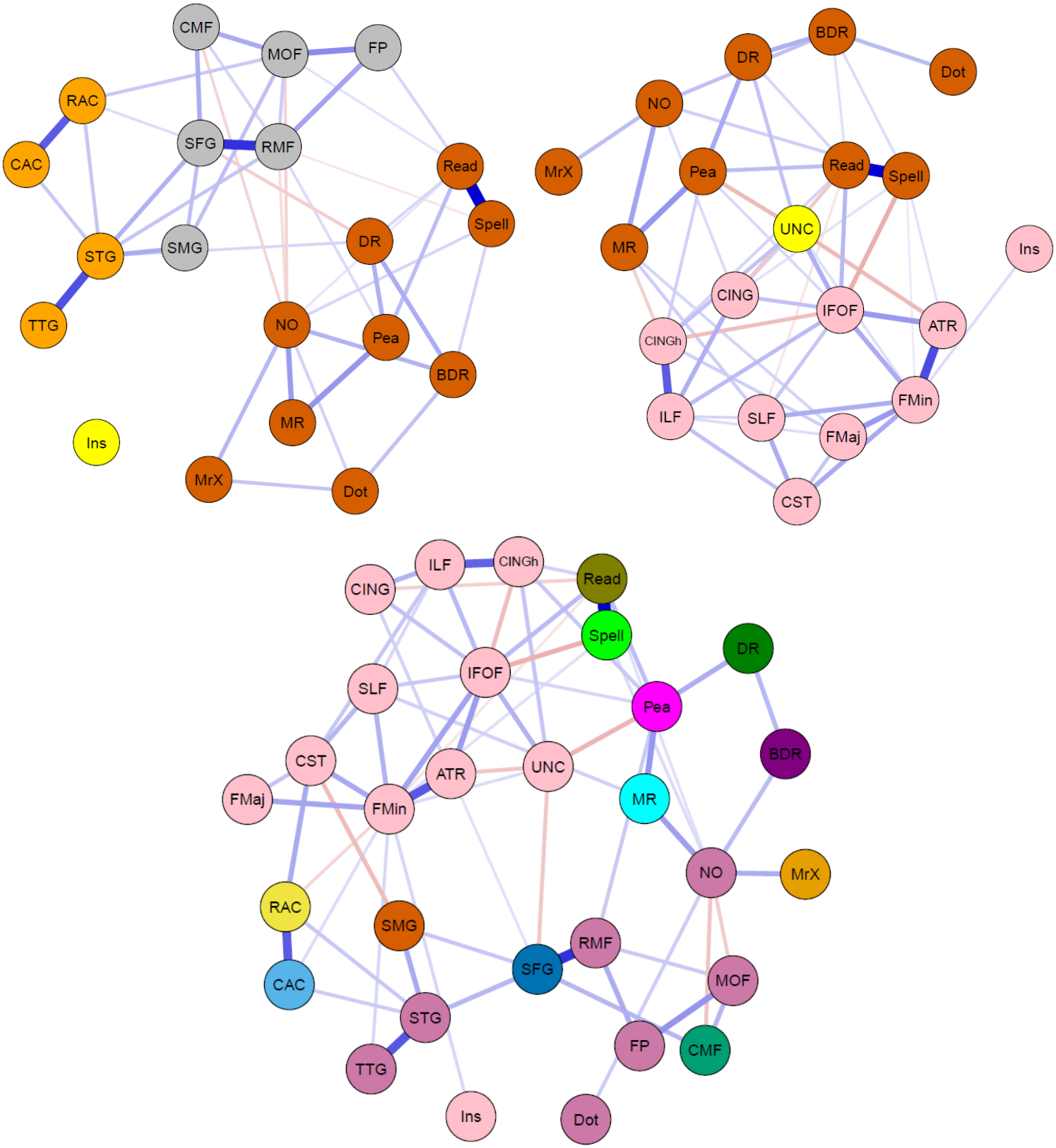


Figure 15. Network visualisations (spring layout) of partial correlation multilayer networks for CALM data. Colors indicate groups determined by the Walktrap algorithm (see above). Top: Bi-layer networks consisting of cognition and grey matter (left), and cognition and white matter (right). Bottom: Tri-layer network consisting of cognition, grey matter and white matter (centre).

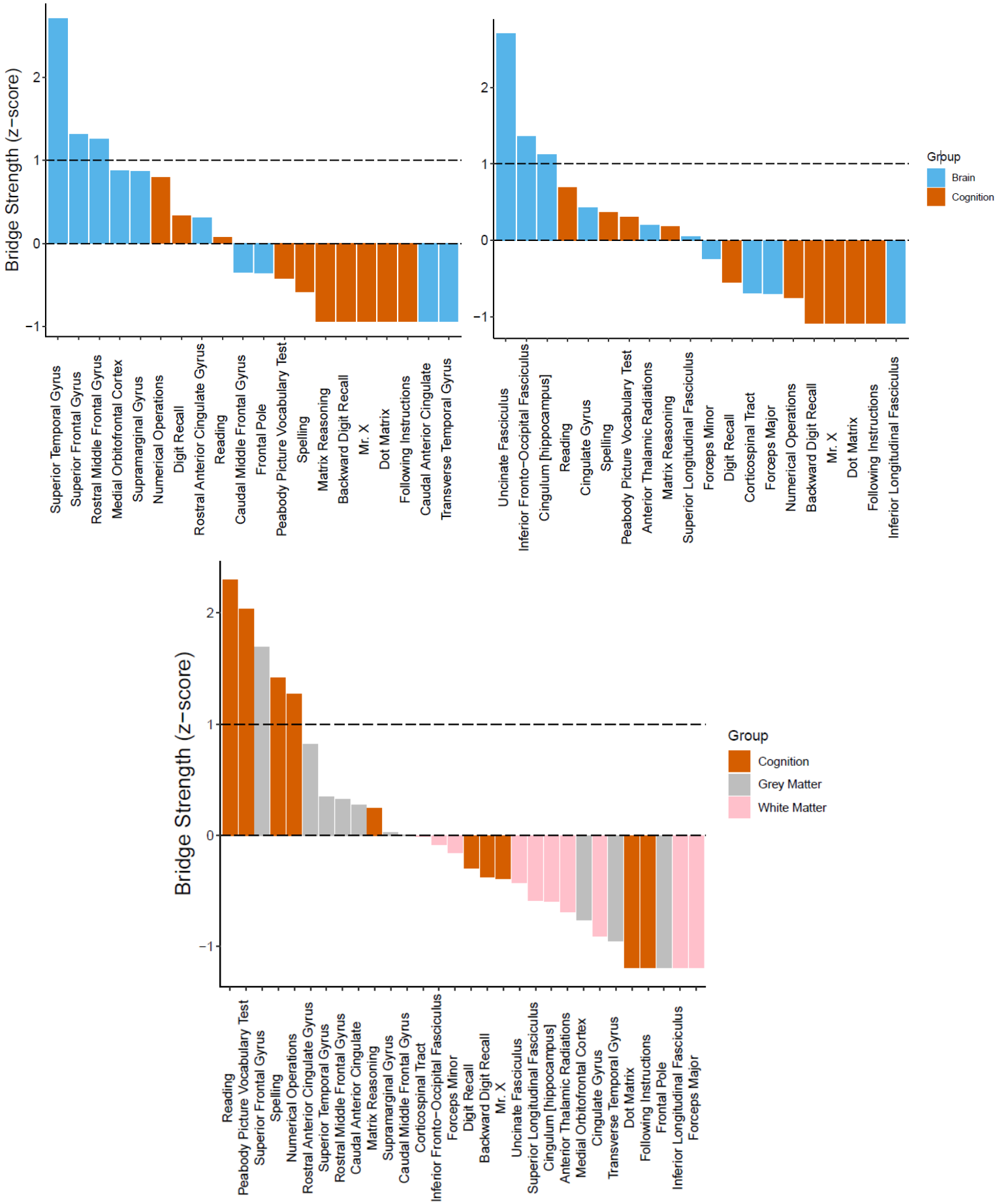

Bridge Strength (z−score)
2
1
0
−1
Superior Temporal Gyrus
Superior Frontal Gyrus
Rostral Middle Frontal Gyrus
Medial Orbitofrontal Cortex
Supramarginal Gyrus
Numerical Operations
Digit Recall
Rostral Anterior Cingulate Gyrus
Reading
Caudal Middle Frontal Gyrus
Frontal Pole
Peabody Picture Vocabulary Test
Spelling
Matrix Reasoning
Backward Digit Recall
Mr. X
Dot Matrix
Following Instructions
Caudal Anterior Cingulate
Transverse Temporal Gyrus
Uncinate Fasciculus
Inferior Fronto−Occipital Fasciculus
Cingulum [hippocampus]
Reading
Cingulate Gyrus
Spelling
Peabody Picture Vocabulary Test
Anterior Thalamic Radiations
Matrix Reasoning
Superior Longitudinal Fasciculus
Forceps Minor
Digit Recall
Corticospinal Tract
Forceps Major
Numerical Operations
Backward Digit Recall
Mr. X
Dot Matrix
Following Instructions
Inferior Longitudinal Fasciculus
Group
Brain
Cognition
Bridge Strength (z−score)
2
1
0
−1
Reading
Peabody Picture Vocabulary Test
Superior Frontal Gyrus
Spelling
Numerical Operations
Rostral Anterior Cingulate Gyrus
Superior Temporal Gyrus
Rostral Middle Frontal Gyrus
Caudal Anterior Cingulate
Matrix Reasoning
Supramarginal Gyrus
Caudal Middle Frontal Gyrus
Corticospinal Tract
Inferior Fronto−Occipital Fasciculus
Forceps Minor
Digit Recall
Backward Digit Recall
Mr. X
Uncinate Fasciculus
Superior Longitudinal Fasciculus
Cingulum [hippocampus]
Anterior Thalamic Radiations
Medial Orbitofrontal Cortex
Cingulate Gyrus
Transverse Temporal Gyrus
Dot Matrix
Following Instructions
Frontal Pole
Inferior Longitudinal Fasciculus
Forceps Major
Group
Cognition
Grey Matter
White Matter

Figure 16. Bridge centrality estimates (z-scores) for multilayer networks. Top: Bi-layer networks consisting of cognition and grey matter (left), and cognition and white matter (right). Bottom: Tri-layer network consisting of cognition, grey matter and white matter (centre). Dashed lines indicate mean strength and one standard deviation above the mean.

Discussion

Summary and interpretation of findings

In this study, I used network analysis (partial correlations) to examine the neurocognitive structure of general intelligence in a childhood and adolescent cohort of struggling learners (CALM). For the single-layer networks (Figure 14), I found that cognitive, grey matter, and white matter networks contained mostly (if not all) positive partial correlations. Moreover, in all single-layer networks, at least two nodes emerged as more central than others (as indexed by node strength equal to or greater than one standard deviation above the mean), which varied in stability from moderately to highly reliable. In the cognitive network, this included two measures of verbal ability (specifically Reading and Peabody Picture Vocabulary Test) and crystallised intelligence (i.e., Numerical Operations). In the structural brain networks (grey matter cortical volume and white matter fractional anisotropy), two nodes each from the grey matter network (superior temporal gyrus and rostral middle frontal gyrus) and white matter network (forceps minor and inferior longitudinal fasciculus) passed the centrality threshold.

Furthermore, I extended previous approaches by integrating a cognitive network with networks of structural brain data, forming bi- and tri-layer networks (Figure 15). Doing so, I observed multiple (both positive and negative) partial correlations between brain and behaviour variables. Using bridge strength as a metric, I found that, in the bi-layer networks, only neural nodes harboured significant connections across communities (defined by the Walktrap algorithm) and levels of organisation (Figure 16, top). In contrast, in the tri-layer network, I found support that mostly cognitive nodes connect across different communities (Figure 16, bottom). Overall, my results suggest that specific behavioural and neural variables serve as 'bridges' between the brain and cognition within general intelligence. These variables may exert greater influence among nodes. Alternatively, these variables might be more influenced by other nodes. However, the literature on drawing inferences from networks to the most likely consequences of intervening on the network is complex and rapidly changing, (e.g., Dablander and Hinne, 2019; Henry et al., 2020; Levine and Leucht, 2016).

Each node in the cognitive network corresponded to a single cognitive task (e.g., Matrix Reasoning) while partial correlations (weighted edges) between nodes were interpreted as compatible with (possible) causal consequences of interactions among cognitive

abilities during development. This interpretation is compatible with the mutualism theory of cognitive development (van der Maas et al., 2006), whereby cognitive abilities positively reinforce each other (e.g., positive partial correlations) over time to produce the positive manifold (Spearman, 1904). Mutualism hypothesises that general intelligence emerges from causal interactions among abilities rather than a general latent factor (Fried, 2020; Kan et al., 2019). Hence, cognition is viewed as a *complex system* derived from the dynamic relations of specific abilities that become more intertwined over development.

The existence of only positive edges in the cognitive network would be expected under a mutualistic perspective (interactions among cognitive variables), which at its essence is a network theory of general intelligence. However, longitudinal analyses are needed to further substantiate this claim. Initially, it was surprising that two of the three most central nodes (i.e., Reading and Peabody Picture Vocabulary Test) relate to verbal ability rather than abilities such as fluid intelligence and working memory (Matrix Reasoning and (forward and backward) Digit Recall), which are traditionally viewed as having strong causal influences on cognitive development (Cattell, 1971). However, an emerging body of literature suggests that verbal ability plays a crucial role in cognitive development (e.g., between reading and working memory before 4th grade, Peng et al., 2018 and Zhang and Malatesha Joshi, 2020), as well as driving the emergence of reasoning (Kievit et al., 2019; also see Gathercole et al., 1999)).

As for the neural networks (here, grey matter cortical volume and white matter fractional anisotropy), individual nodes were comprised of a single ROI. Importantly, I did not interpret weighted edges as an index of direct connectivity. Instead, the presence of strong associations between these ROIs would be compatible with the hypothesis of coordinated development (see Alexander-Bloch et al., 2013), whereby certain brain regions show preferential correlations to each other than more peripheral regions over time (e.g., childhood to late adolescence) as well as the notion of "rich" (Heuvel and Sporns, 2011) and "diverse" (Bertolero et al., 2017) clubs that enable local and global integration. The most central grey matter node was the superior temporal gyrus, which has been implicated in verbal reasoning (e.g., Khundrakpam et al., 2017). Regarding white matter, the two strongest nodes were the forceps minor and inferior longitudinal fasciculus, which while not anatomically close, instead represent long-range connections (see de Mooij et al., 2018) that have been linked to

mathematical ability (Navas-Sánchez et al., 2014) and visuospatial working memory (Krogsrud et al., 2018).

Finally, I integrated both domains (cognitive abilities and brain metrics) into combined *multilayer networks* (cognition-grey matter, cognition-white matter, and cognition-grey and-white matter). Doing so enabled *simultaneous* comparison and *integration* across explanatory levels within the same analytical paradigm (network analysis) and statistical metrics (partial correlations, centrality, and community detection). From this analysis, three observations immediately stood out. First, there were multiple partial correlations between cognitive and neural nodes (especially in the cognitive-white matter and cognitive-grey matter and- white matter networks). Second, compared to the single-layer networks, the multilayer networks have more negative partial correlations. Together, these two findings further suggest that associations between the brain and cognition are complex as they defy straightforward (e.g., only positive and/or one-to-one) relationships and interpretations. However, it should be noted that causality (e.g., conditioning on colliders, see Rohrer, 2018 for overview of interpretations of correlations in graphical causal models in observational data) becomes even more difficult to determine with networks that include multiple levels of organisation (e.g., cognition and structural brain covariance). Finally, I found a peculiar role of the cognitive task Following Instructions (Ins) within all multilayer networks. For example, in the cognitive-grey matter network, Ins had no partial correlations with any other nodes within the network while in both the cognitive-white matter and tri-layer network (cognition, grey and white matter) Ins *only* correlated with the forceps minor (FMin), a neural node, and not any of the cognitive variables. This might suggest that Following Instructions, traditionally a working memory task and often analysed using structural equation modelling, may have distinct psychometric properties (e.g., one-to-one mapping) when compared to other cognitive tasks when modelled through network science approaches, and/or when adjusted for all shared correlations.

Further inspection of bridge strength centrality showed an interesting pattern: (discounting the one standard deviation cut-off) the neural nodes are stronger than the cognitive variables within the multilayer networks, despite there being an equal number of cognitive nodes for each brain metric. This is possibly due to the large number of edges between them (grey and white matter regions) and both cognitive and other neural nodes. In other words, since the neural nodes contain a larger number of connections (partial

correlations) across explanatory levels, they display greater bridge strength (bridge strength sums inter-network correlations).

In other ways, the multilayer networks differed. First, in the tri-layer network, the four of the five central nodes were cognitive variables while, in the bi-layer networks, the central nodes were neural ROIs. Three of these central cognitive nodes in the tri-layer network (Reading, Peabody Picture Vocabulary Test, and Numerical Operations) were also found to be central in the single-layer cognitive network. This further suggests the importance of mathematical and verbal ability in understanding the cognitive neuroscience of general intelligence. Secondly, the fact that cognitive nodes were found to be central only in the tri-layer network suggests that grey and white matter, while related, possibly reveal unique information about cognition when combined and analysed together simultaneously.

However, particularly for the cognition-grey matter bi-layer network, the conclusion that the presence of *only* central neural nodes indicates that these ROIs form strong connections 'between' brain and cognition might be confounded due to the ten cortical volume nodes being clustered into two separate communities. For instance, for the two most central regions in this network, the superior temporal gyrus and the superior frontal gyrus, the majority of partial correlations bridge between the two grey matter clusters rather than the cognitive group. On the other hand, in the cognition-white matter network, the Walktrap algorithm produced a more straightforward grouping (besides UNC) of brain and behaviour. For this network, the three strongest bridge nodes (the uncinate fasciculus, inferior fronto-occipital fasciculus, and hippocampal cingulum) show more associations between neural and cognitive clusters, which support the conclusion of bridging 'between' brain and cognition. Lastly, for the tri-layer network, the large number of clusters identified (i.e., 15 including age) further complicates the interpretations of the findings. Perhaps the clearest example of this is the Peabody Picture Vocabulary Test (Pea), which connects to four separate communities, *three* of which are cognitive nodes. Moreover, the Pea and these three nodes each form communities of which they are the *sole* member. Therefore, before any strong claims can be made, these analyses should be conducted using different cognitive batteries (but similar cognitive domains) to discern whether this clustering is corroborated.

Limitations of the present study

This study contains several limitations that require caution when interpreting the results. First and foremost, these findings are based on cross-sectional data. While adequate to help tease apart individual differences in cognition between people, cross-sectional data cannot be used to elucidate differences in changes within individuals over time, such as during development. Therefore, longitudinal analyses are needed before attempting to make strong inferences about the *dynamics* of these networks. Reiterating this point, a recent study using intelligence data (Schmiedek et al., 2020) found that a cross-sectional analysis of the *g* factor of cognitive ability was unable to capture within-person changes in cognitive abilities over time. This finding further stresses the necessity to integrate cross-sectional (between-person) differences and longitudinal (within-person) changes when studying cognition.

Moreover, as also detailed in Chapter Two, the CALM sample represents an atypical sample (Holmes et al., 2019), with participants who consistently score lower on measures of attention, learning and/or memory than age-matched controls (see Figure 8 (Level I) of Chapter Two for comparison to a typically developing sample). As a result, these analyses would need to be replicated in additional (ideally larger) samples with different cognitive profiles before the results can be said to generalise. This shortcoming of the present study is echoed by the low stability estimates found for the centrality values in the bi-layer networks, which might be due to the differences between the sample sizes of the neural data (grey matter, N = 246; white matter, N = 165) compared to cognition (N = 805). Interestingly, the tri-layer network showed moderate bridge strength stability, but also displayed weak modularity. Moreover, given that the Walktrap algorithm produced 15 communities in the network, which contained only 31 nodes (including age), I further state that this result should be interpreted with caution and must be corroborated in larger cohorts (e.g., ABCD study, Casey et al., 2018).

Summary of the Chapter

This chapter builds on the findings from Chapter Two, by using a network psychometrics approach to understand individual differences in cognitive ability (general intelligence) with structural covariance networks derived from structural brain properties (i.e., grey matter cortical volume and white matter fractional anisotropy). In doing so, I created a *network of networks*, which differs from multiplex (same nodes, different edge types across layers) and multi-slice (same nodes and edge types over time such as in fMRI time-series data) networks (see figure 4.1 of Bianconi, 2018). The advantages of applying this approach are threefold and complementary. First, it places the brain and behaviour, which often do not map onto each other in a simple and reductionist one-to-one fashion, into the same analytical paradigm (network analysis). This allows for *simultaneous* estimation and easier visualisations of potential causal links between cognition and structural brain properties, which to my knowledge, has only been done in a similar way in two other studies, one involving depression (Hilland et al., 2020), the other in autism (Bathelt et al., 2020). Second, it enables the use of community detection algorithms to tease apart major clusters of cognitive abilities, which could help pinpoint potential intervention targets (e.g., using cognitive training and/or transcranial magnetic stimulation). Lastly, it aids in establishing a coherent framework for theory building, which has been lacking in both the neuroscience (Levenstein et al., 2020) and psychological (Fried, 2020) literature, by treating both the brain (algorithmic) and behaviour (computational) as equally important levels of analysis to study (Marr and Poggio, 1976), and attempting to more directly translate findings from one level to the other.

In Chapter Four, I finally overcome a key limitation of the past two empirical chapters: the lack of longitudinal data. In the IMAGEN study (Schumann et al., 2010), I use longitudinal structural equation modelling to examine how brain-behaviour relations change over time from mid-adolescence to early adulthood (ages 14 to 22 years). By doing so I statistically test for evidence in support of the structural scaffolding hypothesis mentioned in Chapter One (Figure 7, also see Kievit and Simpson-Kent, 2021).

Chapter Four

Longitudinal Analyses of Individual Differences and Change between Human Cortical Grey Matter Structure and Measures of Decision-making, Risk-related Behaviours, and Spatial Working Memory from Adolescence to Early Adulthood in the IMAGEN Study

Chapter Four is in preparation for submission to a journal for publication and includes additional co-authors.

Although first-person pronouns are used throughout the chapter, this work is the result of a collaborative research project. My contributions to this study are:

1. I led conceptualisation and planning (aided by R. A. Kievit) about the scientific hypotheses, analysis methods, and interpretations of the project.
2. I performed all manuscript analyses (aided by R. A. Kievit).
3. I wrote the first full draft (with input from R. A. Kievit) of the original manuscript draft.

Introduction

In Chapters Two and Three, I analysed cross-sectional data in two samples (i.e., CALM and NKI-Rockland) using structural equation modelling (SEM) and network analysis, respectively, to tease apart the brain-behaviour relationships between brain structure and measures of intelligence in childhood and adolescence. However, a major limitation of these analyses was the lack of repeat (longitudinal) assessments to not only examine how these variables differ between individuals, but also how they change within individuals over time. Hence, in the final empirical chapter of this thesis, I will describe longitudinal analyses used to investigate the structural scaffolding hypothesis of human intelligence (see Figure 7). Specifically, I use latent growth curve modelling to estimate intercept and slope associations between grey matter structure (i.e., cortical surface area, thickness, and volume) and behavioural performance (here measured by decision-making, risk-related behaviours, and spatial working memory) from mid-adolescence to early adulthood (ages 14 to 22 years), all performed in the IMAGEN study.

Adolescence, defined as the developmental period between childhood and adulthood (e.g., ages 10 and 24 years, Sawyer et al., 2018), is marked by distinct changes in biological (i.e., physical maturation resulting from pubertal hormones), cognitive (e.g., increase in risky behaviours more likely to result in injury or death, see Eaton et al., 2006), and social (e.g., advanced levels of schooling such as high school and undergraduate studies) functioning. Moreover, both cross-sectional and longitudinal evidence has found associations between adolescence and developmental differences in brain structure such as synaptic pruning (Huttenlocher, 1990), white matter microstructure (Simmonds et al., 2014) and volume (Vijayakumar et al., 2018), and grey matter metrics such as cortical surface area, thickness, and volume (Becht and Mills, 2020; Ducharme et al., 2015; Giedd et al., 1999; Tamnes et al., 2017; Vijayakumar et al., 2016).

Furthermore, differences in adolescent brain structure are related to various cognitive and affective functions ranging from decision-making and risk-related behaviours (Steinberg, 2008) and social cognition (i.e., the social brain, see Blakemore, 2008 and Andrews et al., 2021) to intelligence measures (Ferrer, 2018; Ritchie et al., 2019; Wendelken et al., 2017). For example, in a previous release of the IMAGEN cohort (also used here, see Schumann et al., 2010), Ritchie et al., 2019 investigated baseline and rate of change (slope) relationships

between a general factor of cognitive ability (obtained by extracting the first component of a battery of CANTAB tasks) and measures of cortical brain structure (i.e., total surface area, mean cortical thickness, and grey matter volume). Using latent change score modelling (Kievit et al., 2018a), they observed that the general cognitive ability factor (based on CANTAB tasks, see Haring et al., 2015) at baseline (age 14 years) had significant associations with surface and volume, but not cortical thickness. Moreover, higher scores on the baseline CANTAB factor predicted more rapid cortical thinning and volume loss between 14 and 19 years. In contrast, baseline brain structure did not significantly correlate with change/slope of the general CANTAB factor.

A recent review of developmental associations between brain structure and intelligence from childhood to early adulthood found that longitudinal studies (i.e., at least two time points) that combine both brain structure and intelligence-related measures are scarce, with about 30 total studies to date (see Table 7.1 and Figure 7.2 of Kievit and Simpson-Kent, 2021). Furthermore, most of these studies have been published within the last five years, indicating that this approach is very much in its infancy. Although only a modest number of studies exist, several patterns emerged: In multiple studies, baseline brain structure (e.g., grey matter cortical thickness and white matter fractional anisotropy) significantly predicted the rate of change in cognitive performance over time (Ferrer, 2018; Wendelken et al., 2017). In other words, individuals with 'better' brain structure such as greater white matter integrity on average also showed larger gains (in children and adolescents) than individuals with lower scores on structural brain imaging metrics. This pattern, termed 'structural scaffolding', suggests that the structural integrity of the brain at a previous time point lays the foundations for later changes in cognition over time. On the other hand, it was also found that baseline cognitive scores significantly predict changes in brain structure over time (e.g., Ritchie et al., 2019). This is referred to as 'adaptive reorganisation'. Lastly, 'correlated change' could also exist between the slopes of behavioural performance and brain structure such that changes in one of these measures are correlated.

The purpose of the present study was to further investigate structural scaffolding, as well as adaptive reorganisation and correlated change from adolescence to early adulthood (ages 14 – 22 years). In the IMAGEN consortium, I fit latent growth curve models (Duncan and Duncan, 2004) to estimate the baseline and longitudinal associations between measures

of three psychometric domains (decision-making, risk-related behaviours, and spatial working memory), and three metrics of cortical brain structure (i.e., surface area, thickness, and volume). In contrast to my hypothesis, I found evidence for adaptive reorganisation but not structural scaffolding. Furthermore, findings suggested that there were no correlated changes between behavioural and brain structure slopes.

Methods

Participants: The IMAGEN study

Participants from the present study were obtained from the IMAGEN study (Schumann et al., 2010), a longitudinal (four time points, three of which include neuroimaging and are analysed in this study), multi-centre (research sites are located in Berlin, Dresden, Dublin, Hamburg, London, Mannheim, Nottingham, and Paris) consortium formed to study reinforcement-related behaviour in both typical and atypical brain function associated with psychopathology. For this study, I included participants across three time points (waves) from middle adolescence to early adulthood (wave 1 ≈ 14 years old, N = 2072; wave 2 ≈ 19 years old, N = 1436; wave 3 ≈ 22 years old, N = 1256). I did not include the first follow up after baseline in my analyses because neuroimaging data were not collected during this assessment.

Cognitive assessments:

Decision-making, risk-related behaviours, and spatial working memory

Previous studies tend to focus on more narrow cognitive domains. However, this approach might fail to discover broader behavioural patterns of change, as well as specific associations between behavioural domains (see Table 4) and different brain metrics (Figure 18). Therefore, here I examine a broader sample of eight measures, which I group descriptively into three behavioural domains. First, as representative of classic intelligence-type tasks, I used two spatial working memory measures (Between Errors and Strategy). Beyond standard general cognitive ability measures, I also incorporated measures of decision-making (Delay Aversion, Deliberation Time, and Quality of Decision-making), and risk-related behaviours (Overall Proportion Bet, Risk Adjustment, and Risk Taking). These will allow me to compare and contrast the neurocognitive development of intelligence and other domains. See Table 4 for descriptions and summary statistics of behavioural measures for all three waves.

Behavioural measures originated from the from the Cambridge Neuropsychological Test Automated Battery (CANTAB, Haring et al., 2015), which includes spatial working memory and the Cambridge Gambling Task (CGT, renamed Cambridge Guessing Task for the IMAGEN study, Deakin et al., 2004; Rogers et al., 1999), which contained the measures of decision-making and risk-related behaviours.

During the spatial working memory portion of CANTAB, participants are presented with varying numbers of coloured boxes shown on a screen. The goal of this task is to (via process of elimination) find the boxes that contain a yellow token inside them and use those tokens to fill up an empty column positioned on the right-hand side of the screen. Difficulty (more boxes to choose from) increases until up to 12 boxes at a time are displayed for participants to search and pick from. Over successive trails, both the colour and location of the boxes are changed to inhibit stereotyped search strategies used for selection.

For the Cambridge Gambling/Guessing Task (CGT), participants are presented with a row of ten boxes (located at the top of a screen), with some coloured in blue while the others are coloured in red. One of these boxes contains a yellow token, which is concealed inside. Across trials, the ratio of red to blue boxes changes but, in every trial, one box (either blue or red) always contains a yellow token. In each trial, participants select the box colour they think contains the token. Participants are also given 100 points at the start of the task, a proportion of which they bet (by pressing a button when the value displayed on the screen matches the amount they wish to bet) per decision on each trial. The bet value counter increases (ascending) or decreases (descending) depending on the trial and points are added or subtracted for correct (yellow token present in box) or incorrect (wrong box colour choice) responses, respectively.

Notably, for both spatial working memory measures, Between Errors and Strategy, and Delay Aversion (decision-making), higher scores designate poorer performance. For the other tasks, higher scores either indicate better performance (i.e., Quality of Decision-making) or their interpretations are ambiguous regarding the magnitude of the score (i.e., Deliberation Time, Overall Proportion Bet, Risk Adjustment, and Risk Taking). See Figure 17 for age trends of the behavioural variables analysed as a part of this study.

Behavioural Domain	Measure Descriptions	Mean (sd) [range]	Missing Data	Task
Decision-making	**Delay Aversion:** difference (ascending vs descending) in proportion a participant bet across trial conditions.	Wave 1: 0.24 (0.14) [-0.20, 0.70] Wave 2: 0.19 (0.14) [-0.13, 0.65] Wave 3: 0.16 (0.13) [-0.20, 0.60]	Wave 1: 18.82% Wave 2: 34.26% Wave 3: 42.65%	CGT

	Deliberation Time: average time taken by a participant to choose a box colour (red or blue).	Wave 1: 2.00 (0.56) [0.74, 4.15] Wave 2: 1.60 (0.42) [0.73, 3.03] Wave 3: 1.52 (0.43) [0.74, 3.67]	Wave 1: 19.66% Wave 2: 35.19% Wave 3: 42.51%	CGT
	Quality of Decision-making: average proportion of trials whereby a participant chose the box colour (red or blue) containing the yellow token (correct response).	Wave 1: 0.94 (0.08) [0.66, 1.00] Wave 2: 0.96 (0.05) [0.75, 1.00] Wave 3: 0.96 (0.06) [0.55, 1.00]	Wave 1: 20.12% Wave 2: 34.77% Wave 3: 41.77%	CGT
Risk-related Behaviours	**Overall Proportion Bet:** average proportion a participant bets across all trials.	Wave 1: 0.50 (0.13) [0.09, 0.89] Wave 2: 0.49 (0.12) [0.14, 0.83] Wave 3: 0.52 (0.12) [0.11, 0.91]	Wave 1: 18.59% Wave 2: 33.43% Wave 3: 41.77%	CGT
	Risk Adjustment: degree to which participant betting behaviour was influenced by the box ratio (red vs blue). Hence, this measure gauges a participant's tendency to make higher or lower bets based on which box colour is likely to be favourable (contains yellow token) or unfavourable (no yellow token), respectively, on a given trial.	Wave 1: 1.52 (0.96) [-1.02, 4.42] Wave 2: 1.90 (0.98) [-1.04, 4.77] Wave 3: 1.99 (1.00) [-0.82, 4.83]	Wave 1: 18.64% Wave 2: 33.47% Wave 3: 41.77%	CGT
	Risk Taking: average proportion of total points a participant bets on trials when they chose the most likely outcome.	Wave 1: 0.54 (0.14) [0.13, 0.95] Wave 2: 0.54 (0.12) [0.17, 0.90] Wave 3: 0.57 (0.13) [0.11, 0.90]	Wave 1: 18.54% Wave 2: 33.43% Wave 3: 41.77%	CGT
Spatial Working Memory	**Between Errors:** a measure of the tendency (i.e., number of times) of participants to go back to boxes previously found to contain a yellow token.	Wave 1: 19.40 (13.64) [0, 62] Wave 2: 10.46 (10.12) [0, 47] Wave 3: 11.75 (10.94) [0, 49]	Wave 1: 4.64% Wave 2: 53.69% Wave 3: 42.70%	CANTAB

	Strategy: quantifies how often (i.e., number of times) participants began a new search by selecting a different box and, therefore, deviating from a more efficient 'predetermined sequence' of revisiting a box where a yellow token was previously found.	Wave 1: 31.28 (5.43) [18, 46] Wave 2: 27.66 (6.20) [18, 45] Wave 3: 27.74 (6.13) [10, 46]	Wave 1: 3.94% Wave 2: 52.67% Wave 3: 41.86%	CANTAB

Table 4. Behavioural domains, descriptions, summary statistics, missingness, and test battery of IMAGEN behavioural data across all three waves (ages 14 – 22 years) used in this study. Descriptions of measures are paraphrased from Cacciamani et al., 2018 and Flouri et al., 2019.

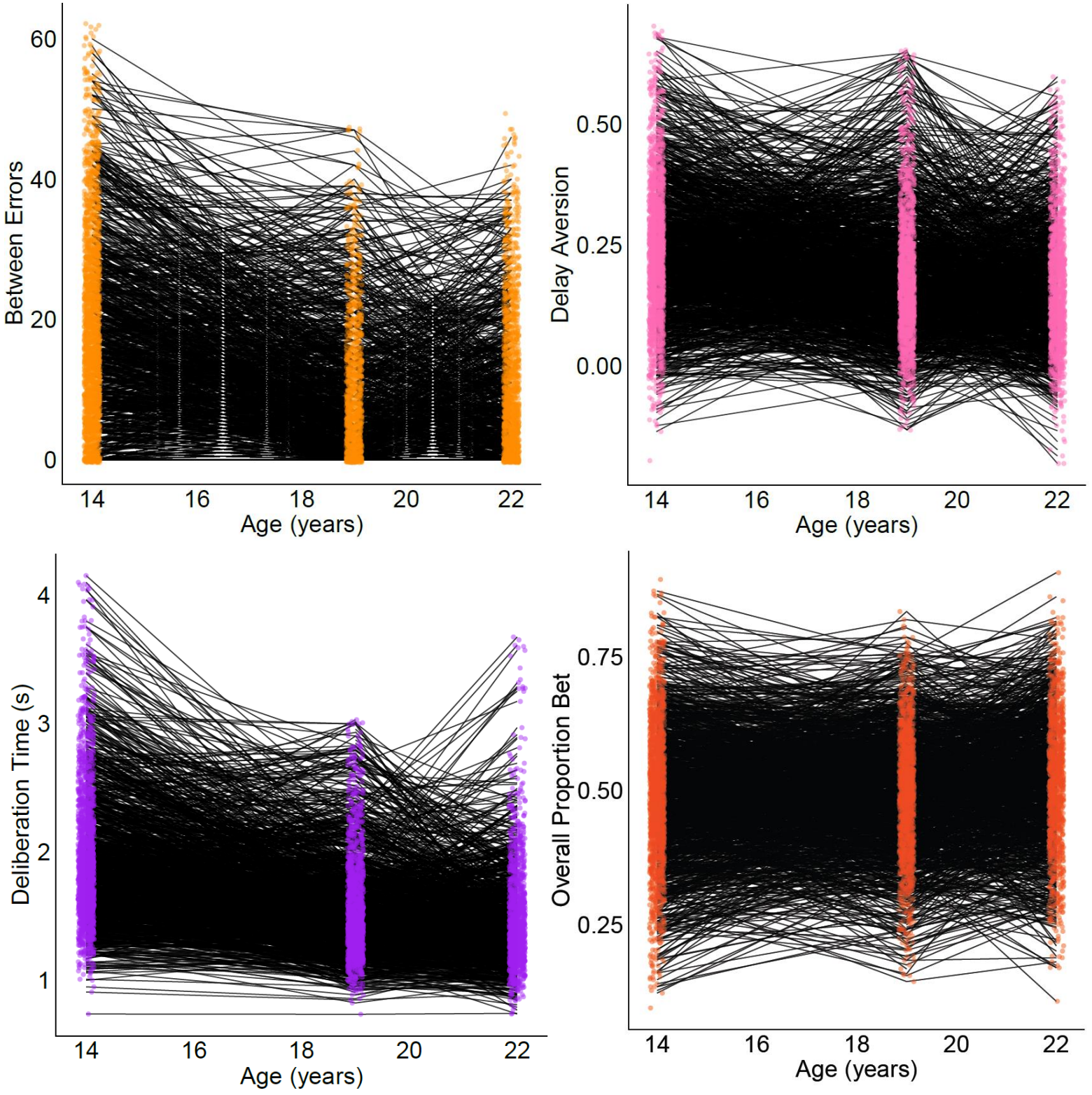


Figure 17 Top. Spaghetti plots showing (first half of) age trends of IMAGEN behavioural data across all three waves (ages 14 – 22 years).

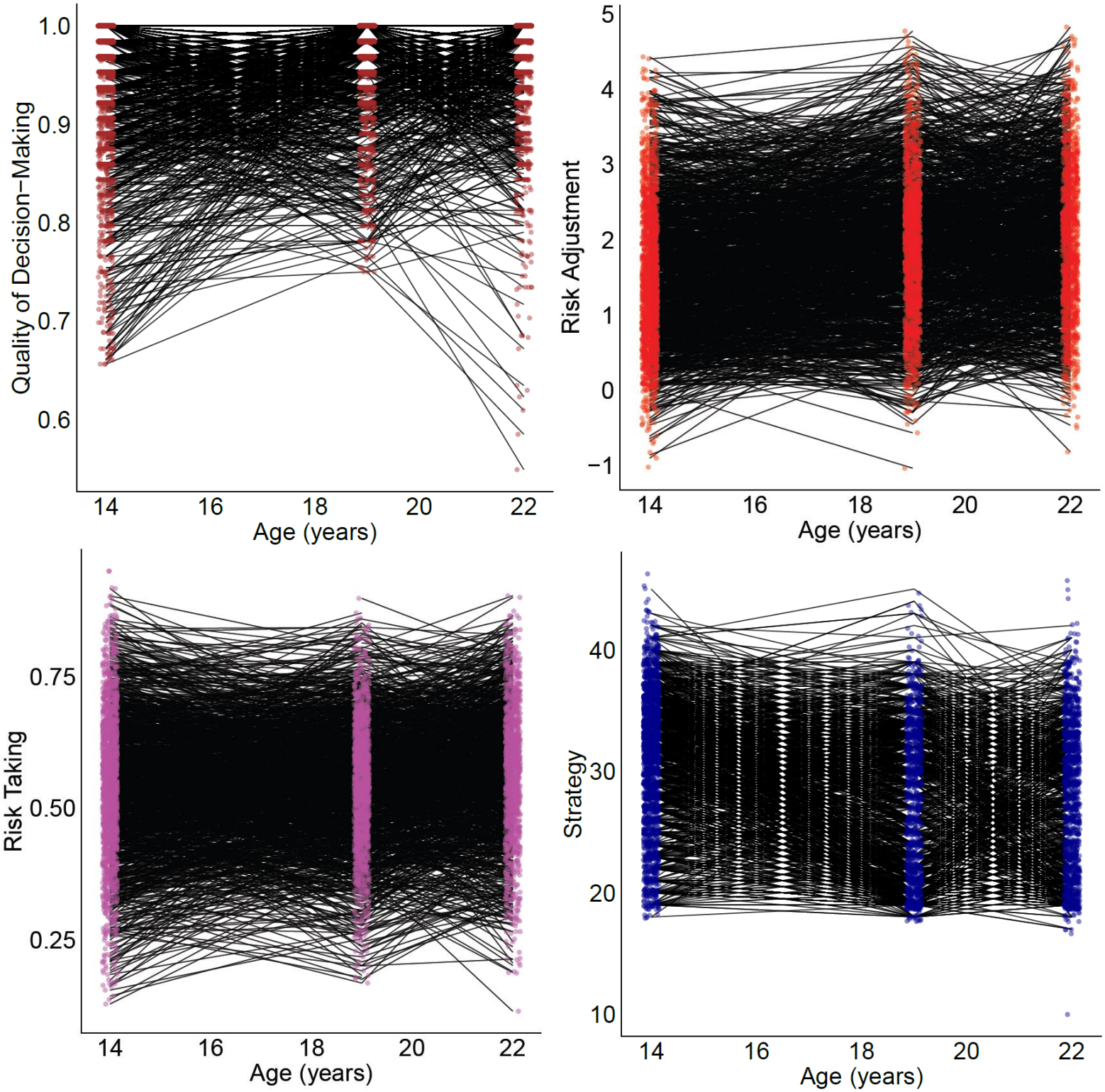


Figure 17 Bottom. Spaghetti plots showing (second half of) age trends of IMAGEN behavioural data across all three waves (ages 14 – 22 years).

Structural MRI measures: Cortical surface area, thickness, and volume

The IMAGEN study neuroimaging data were obtained on the 3T MRI scanners from four different manufacturers (Bruker, Ettlingen, Germany; General Electrics, Chalfont St Giles, UK; Philips, Best, The Netherlands; and Siemens, Munich, Germany) at eight different European locations (i.e., Berlin, Dresden, Dublin, Hamburg, London, Mannheim, Nottingham, and Paris). Scanning protocol parameters, particularly those related to image contrast or signal-to-noise ratio were synchronised (held constant) across manufacturers at all acquisition sites. The 3D T1-weighted images were acquired using a 3D magnetisation-prepared rapid acquisition gradient echo (MPRAGE) sequence with 1.1 millimetre (mm) isotropic image resolution, and is based on the ADNI protocol (see Jack et al., 2008) with the following parameters: Repetition Time (TR) = 2300 milliseconds (ms); Echo Time (TE) = 2.80 ms; Inversion Time (TI) = 900 ms; flip angle = 8 – 9 degrees; voxel dimensions = 1.1 mm isotropic. All segmentation of structural MRI data was performed with FreeSurfer version 5.3.0 (see http://surfer.nmr.mgh.harvard.edu). For details about quality control measures, see Supplementary Tables 3 and 4 of Schumann et al., 2010.

The neural data used in this study consisted of three cortical grey matter metrics: cortical surface area, thickness, and volume. Regions-of-interest (ROIs) were derived from the Desikan-Killiany atlas (Desikan et al., 2006). Rather than incorporating whole-brain grey matter cortical structure into the analyses, I instead selected a subset of 10 cortical regions (see top Figure 12 from Chapter Three for visualisation) for this study (caudal anterior cingulate, caudal middle frontal gyrus, frontal pole, medial orbitofrontal cortex, rostral anterior cingulate gyrus, rostral middle frontal gyrus, superior frontal gyrus, superior temporal gyrus, supramarginal gyrus, and transverse temporal gyrus). These ROIs were selected based on a meta-analysis on functional and structural correlates of intelligence (Basten et al., 2015), as well as from previous studies where evidence was found for an important role of these regions in predicting and/or correlating with performance on tests of intelligence (Kievit et al., 2018b; also see Chapter Three). I then calculated the average for these 10 ROIs across both hemispheres to create a global mean bilateral estimate of grey matter cortical structure in this network of regions (referred to from here on as Surface Area, Thickness, and Volume). I speculated that, given these 10 regions have been found to robustly correlate with intelligence, they may also play a role in other abilities involved in behaviour/cognition such as decision-

making, risk-related behaviours, and spatial working memory. See Figure 18 for age trends of the neural measures used in this study.

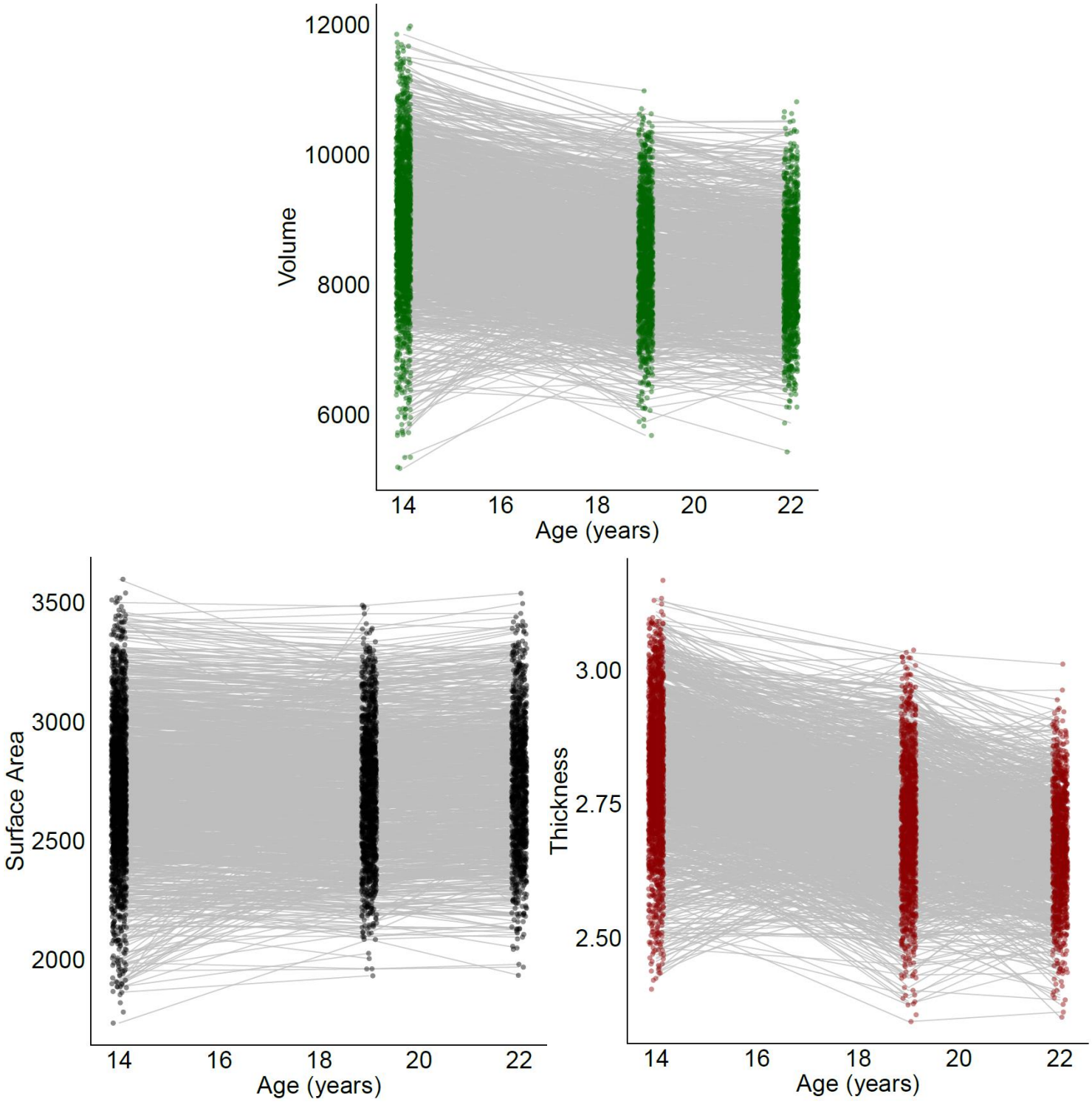


Figure 18. Spaghetti plots showing age trends of IMAGEN neural (grey matter structure) data across all three waves (ages 14 – 22 years).

Statistical analyses: Latent growth curve models

I used latent growth curve modelling (Duncan and Duncan, 2004), a longitudinal class of structural equation modelling techniques (Kline, 2015), which allows the estimation of trajectories (i.e., intercepts and slopes) and associations (i.e., variances and covariances) of the behavioural and brain structure measures across the three time points (age 14, 19, and 22 years). Since, by definition, unobserved/latent variables (LVs) cannot be observed directly, the statistical properties (i.e., intercepts, slopes, and variances) of psychological and neural LVs are estimated through observed/manifest variables such as behavioural and brain structure measures.

To estimate the latent growth curve models, I used the lavaan package (version 0.6-6, see Rosseel, 2012) of the R language (R Core Team, 2020). I implemented two variants of these models: linear (Figure 19 left) and basis (Figure 19 right) models. In the linear model, change (slope) is specified to increase in equal amounts (here, 0.25 per year) for every year between time points/waves of data (t2 – t1 = 5 years; t3 – t2 = 3 years). In contrast, the basis model assumes that change in slope is nonlinear. In terms of model specification, in the basis model the second wave (age 19) is freely estimated (instead of being set to 1.25), with the first and third wave being constrained to 0 and 2, respectively, as in the linear model. If, based on a likelihood ratio test, the basis model outperforms the linear model, it would suggest that, from mid-adolescence to early adulthood (ages 14 – 22 years), participants changed/increased at a rate faster, or more slowly, than would be expected if the growth was equally distributed at each year.

Prior to fitting models, I removed outliers (defined as ± 3 standard deviations from the mean) for both behavioural and neural data. I repeated this procedure for Deliberation Time after noticing additional outliers when plotting the data. To account for data missingness and deviations from multivariate normality, I used robust full information maximum likelihood estimator (FIML, see Enders, 2001). I assessed overall model fit using the robust estimates of chi-squared test, the comparative fit index (CFI), the root mean square error of approximation (RMSEA) with its confidence interval, and the standardised root mean squared residuals (SRMR) (Schermelleh-Engel et al., 2003). Interpretation of model fit was defined as: CFI (acceptable fit 0.95 – 0.97, good fit > 0.97), RMSEA (acceptable fit 0.05 – 0.08, good fit < 0.05), and SRMR (acceptable fit 0.05 – 0.10, good fit < 0.05). Moreover, I compared models (linear vs

basis) using the likelihood ratio test, which determines whether the complexity of the basis model leads to a sufficient improvement in fit. Additionally, I examined the Akaike information criterion (AIC, see Akaike, 1998) and Bayesian information criterion (BIC, see Schwarz, 1978).

For all models I allowed the residual variance of time point 2 (around age 19) to be unique (whereas time points 1 and 3 were constrained to be equal) to achieve proper solutions. Furthermore, fitting some univariate models led to convergence issues (e.g., negative latent variable slope variances). For these models, I incorporated necessary model restrictions (e.g., setting slope to be positive) to achieve proper solutions (e.g., no Heywood cases). Lastly, the best fitting models from the univariate model selection were used in the bivariate models (i.e., brain and behaviour combined). For the bivariate models, I also incorporated model restrictions only where necessary (as with the univariate models). All these modifications were made solely to achieve proper fit, rather than based on other criteria (e.g., parameters of interest).

I evaluated five hypotheses using latent growth curve modelling:

1. Participants will change (i.e., increase or have a non-zero slope) in decision-making, risk-related behaviour, spatial working memory, and cortical brain structure from mid-adolescence (age 14) to early adulthood (age 22).
2. These changes will not be fully linear (i.e., a basis model will outperform a linear model).
3. There will be significant ($p < .05$) individual differences among participants in mean and variances of baseline (intercept) and change (slope) estimates.
4. Intercepts and slopes within the same variables will be significantly correlated. This will be the case for both behavioural and brain variables. The direction of this association was not specified before analyses were conducted.
5. There will be ubiquitous covariance between brain structure slope (e.g., Surface Area, Thickness, and/or Volume) and change in behavioural performance, providing further longitudinal evidence for the structural scaffolding hypothesis (Figure 20).

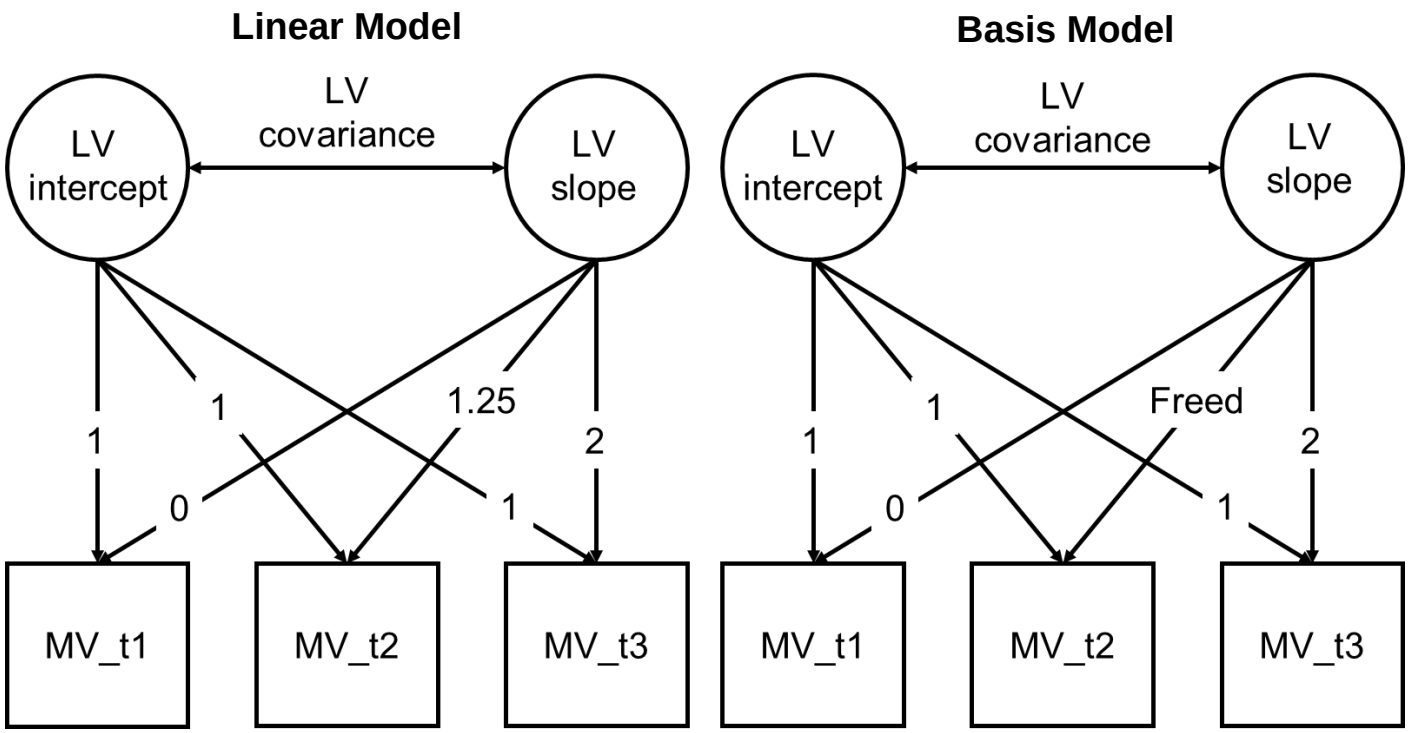


Figure 19. Left: Example of linear univariate latent growth curve model. Right: Example of basis/nonlinear univariate latent growth curve model. Circles and squares represent latent (LV) and manifest (MV) variables, respectively. Single and double-headed arrows indicate factor loadings (intercepts and slopes) and covariances, respectively. Residual variances are also estimated in these models but are not shown for easier illustration.

Bivariate Latent Growth Curve Model: Longitudinal Changes in Brain-Behaviour

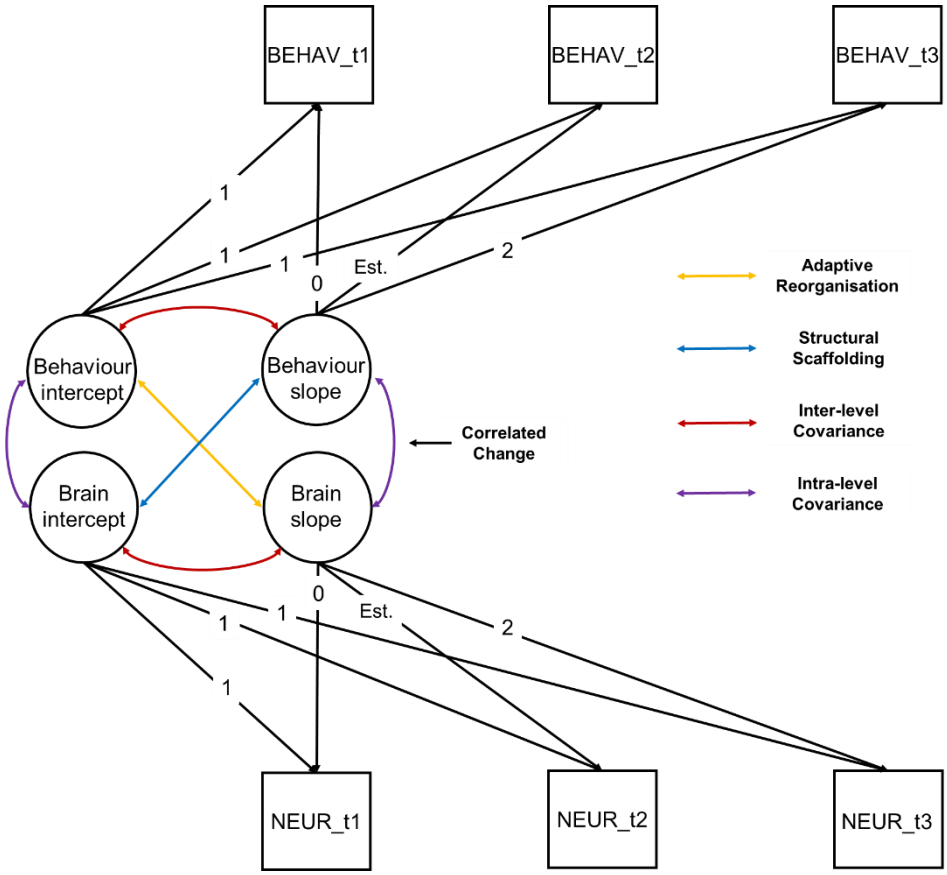


Figure 20. Example of bivariate latent growth curve model illustrating theorised longitudinal changes in brain-behaviour relationships (adaptive reorganisation, structural scaffolding, and correlated change) across development. Est. = slope estimate at time point/wave 2 (1.25 for linear model, freely estimated for basis model). Residual variances are also estimated in these models but are not shown for easier illustration.

Results

Univariate models show that most behavioural and all cortical grey matter structural measures change nonlinearly from mid-adolescence to early adulthood

First, I fit univariate latent growth curve models, separately for each behavioural measure and grey matter metric. For each behavioural domain and brain metric, I compared a linear growth model with a basis (nonlinear) model, where the slope factor loading of the middle time point (i.e., age 19) is freely estimated. If developmental change is truly linear, the penalty for the complexity (estimated using the AIC and BIC) of the basis model will indicate that the linear model is preferred. If, however, change is nonlinear, such that change is more (or less) rapid between ages 14 and 19 than between 19 and 22 (or vice versa), then the basis model will be preferred.

I found that the basis model outperformed the linear model for most behavioural measures (Between Errors: $\chi^2\Delta = 87.16$, $df\Delta = 1$, $AIC\Delta = 179.20$, $BIC\Delta = 173.50$, $p < .001$; Delay Aversion: $\chi^2\Delta = 5.90$, $df\Delta = 1$, $AIC\Delta = 3.54$, $BIC\Delta = 2.08$ (Note: the linear model had a lower BIC, suggesting a small benefit of the linear model), $p < .05$; Deliberation Time: $\chi^2\Delta = 63.06$, $df\Delta = 1$, $AIC\Delta = 89.15$, $BIC\Delta = 83.53$, $p < .001$; Quality of Decision-making: $\chi^2\Delta = 29.26$, $df\Delta = 1$, $AIC\Delta = 81.60$, $BIC\Delta = 76.00$, $p < .001$; Risk Adjustment: $\chi^2\Delta = 10.18$, $df\Delta = 1$, $AIC\Delta = 10.50$, $BIC\Delta = 4.90$, $p < .05$; Strategy ($\chi^2\Delta = 31.42$, $df\Delta = 1$, $AIC\Delta = 40.40$, $BIC\Delta = 34.70$, $p < .001$). All nonlinear models had good fit (Table 5). Exceptions to this trend of nonlinear change in behaviour included Overall Proportion Bet ($\chi^2\Delta = 1.12$, $df\Delta = 1$, $AIC\Delta = 0.59$, $BIC\Delta = 6.22$, $p > .05$; I chose the linear model as the winning model since it had better overall fit, and a lower AIC and BIC), and Risk Taking ($\chi^2\Delta = 2.07$, $df\Delta = 1$, $AIC\Delta = 0.59$ (the basis model had a lower AIC), $BIC\Delta = 5.04$ (the linear model had a lower BIC), $p > .05$; I chose the linear model as the winning model since it had better overall fit, and a lower BIC). All these models also fit well to the behavioural data (Table 5).

Behavioural slope estimates (see Figure 21 left) showed substantial evidence for change during adolescence (i.e., from age 14 until age 22, with all slopes showing a significant mean intercept (all $p < .001$)). Notably, risk-related behaviours (here measured by Overall Proportion Bet, Risk Adjustment, and Risk Taking) was the only construct where the model fit revealed mixed results in relation to the shape of their trajectories (linear or nonlinear). For instance, Risk Adjustment was the only risk-related measure that exhibited nonlinear slope

increases from the first to third wave, while Overall Proportion Bet and Risk Taking grew linearly. Between Errors displayed the largest increase between waves 1 and 2 of any behavioural measure (slope factor loading estimate at wave 2 (age 19) = 2.28; second highest estimate: Quality of Decision-making = 1.85, also at age 19; third highest estimate at age 19: Strategy = 1.84).

Together, these findings overall support hypothesis 1 that behavioural performance in decision-making, risk-related behaviours, and spatial working memory changed (particularly increased) from mid-adolescence (age 14) to early adulthood (age 22). Moreover, these changes in behaviour tended to be nonlinear (supporting hypothesis 2), with exceptions for two risk-related behaviours (Overall Proportion Bet and Risk Taking). This suggests that adolescents change linearly in managing risk as they age into early adulthood.

In terms of cortical grey matter structure, the pattern of change was consistent (Table 5 and Figure 21 right): the basis model outperformed the linear model for each brain metric (Surface Area: $\chi^2\Delta = 30.36$, $df\Delta = 1$, $AIC\Delta = 183.71$, $BIC\Delta = 178.06$, $p < .001$; Thickness: $\chi^2\Delta = 63.23$, $df\Delta = 1$, $AIC\Delta = 60.44$, $BIC\Delta = 54.79$, $p < .001$; Volume: $\chi^2\Delta = 83.77$, $df\Delta = 1$, $AIC\Delta = 212.70$, $BIC\Delta = 207.10$, $p < .001$). This indicated that brain maturation changed nonlinearly (i.e., decreased except for Surface Area, which increased over time, see Figure 18) from mid-adolescence to early adulthood, thereby further supporting hypotheses 1 and 2, this time for neural measures.

Measure	Model	χ^2	df	CFI	RMSEA	SRMR	AIC	BIC
Between Errors	Linear***	121.16	2	.739	.205 [.175 .237]	.127	32971.790	33011.401
	Basis	.66	1	1.000	.000 [.000 .000]	.008	32792.618	32837.888
Delay Aversion	Linear**	10.10	2	.965	.045 [.021 .075]	.026	-5091.057	-5051.699
	Basis*	4.35	1	.984	.043 [.009 .088]	.018	-5094.599	-5049.618
Deliberation Time	Linear***	76.24	2	.823	.157 [.128 .188]	.078	5358.397	5397.732
	Basis**	9.15	1	.982	.071 [.034 .115]	.027	5269.249	5314.203
Overall Proportion Bet	**Linear*****	32.09	2	.954	.089 [.064 .118]	.038	-6574.895	-6535.516
	Basis***	36.97	1	.954	.126 [.091 .164]†	.037	-6574.309	-6529.305

Quality of Decision-making	Linear***	45.45	2	.883	.141 [.107 .178]	.061	-12234.723	-12195.412
	Basis	.76	1	1.000	.000 [.000 .000]	.006	-12316.315	-12271.388
Risk Adjustment	Linear**	12.30	2	.988	.053 [.027 .082]	.022	11515.535	11554.904
	Basis	1.01	1	1.000	.000 [.000 .058]†	.006	11505.007	11550.000
Risk Taking	**Linear*****	33.30	2	.956	.092 [.066 .120]	.038	-6117.493	-6078.120
	Basis***	35.76	1	.957	.127 [.093 .165]†	.037	-6118.075	-6073.078
Strategy	Linear***	40.10	2	.945	.099 [.074 .127]	.051	26912.742	26952.390
	Basis	1.65	1	.999	.013 [.000 .061]†	.010	26872.345	26917.656
Surface Area	Linear***	78.08	2	.939	.241 [.197 .289]	.043	-2151.555	-2112.014
	Basis***	211.71	1	.984	.175 [.140 .212]†	.020	-2335.264	-2290.074
Thickness	Linear***	78.53	2	.914	.134 [.109 .160]	.060	-7071.839	-7032.242
	Basis***	15.26	1	.984	.081 [.049 .120]†	.027	-7132.284	-7087.029
Volume	Linear***	133.11	2	.891	.240 [.207 .276]	.083	10442.303	10481.877
	Basis***	28.27	1	.986	.122 [.086 .163]	.032	10229.578	10274.806

Table 5. Goodness-of-fit indices for linear and basis models of behavioural measures and brain structure (grey matter) metrics. χ^2 = Chi-squared, df = degrees of freedom, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardised Root Mean Residual, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, **denotes the best-fitting model**, * $p < .05$, ** $p < .01$, *** $p < .001$., † indicates that robust estimate could not be properly calculated.

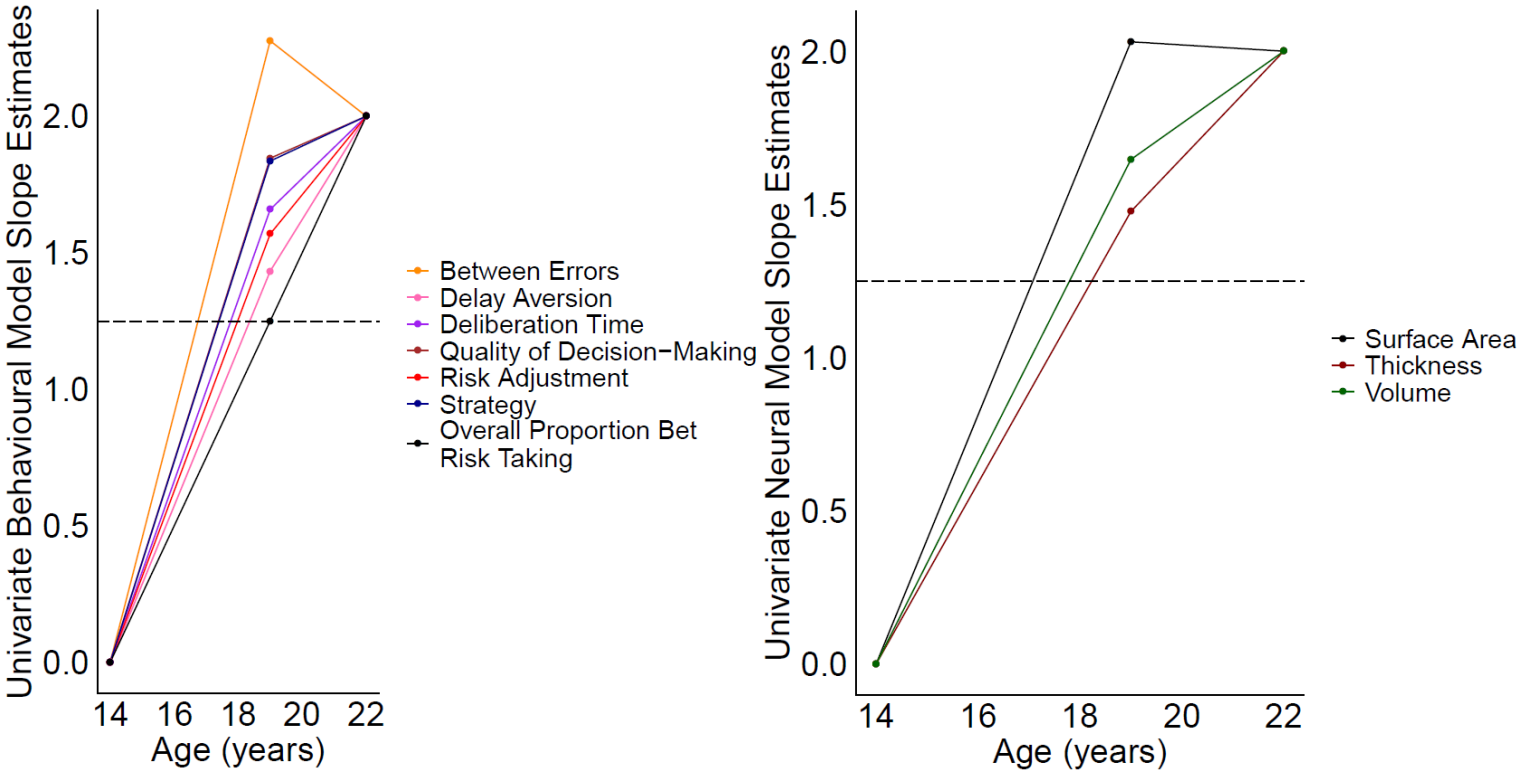


Figure 21. Left: Slope parameter estimates for the winning (linear or basis) univariate behavioural latent growth curve models. Note, Quality of Decision-making and Strategy show substantial overlap in this plot since their slope estimates at age 19 are 1.846 and 1.836, respectively. Right: Slope parameter estimates for the univariate brain structure (grey matter) latent growth curve models, where the basis model was always preferred. Horizontal dashed lines indicate the slope estimate (1.25) for linear growth at wave 2 (age 19).

Univariate model estimates show that participants significantly differed in means and variances for intercepts (baseline) and slopes (rates of change)

Next, I inspected the means and variances for the intercepts and slopes of the univariate models. This revealed (see Table 6) that participants significantly (all $p < .001$ except for Strategy slope variance) differed from each other at both baseline levels (intercepts, age 14 years) and rate of change/slopes (ages 14 – 22 years). This occurred for both the means (except for Surface Area mean slope, $p > .05$) and variances, indicating that individuals not only began assessment at different behavioural and neural levels, but they also showed patterns of change that varied from each other across waves. Lastly, mean intercept-slope covariances showed consistently significant (besides Strategy) and negative correlations, suggesting that, the lower participants 'started off' in behavioural performance or the larger initial brain structure, the more they 'gained' (regardless of whether change specified an improvement or a decline)

during development (from 14 to 22 years). This pattern is commonly found and taken to reflect a ceiling or floor effect of the nearing end of maturation, as well as regression to the mean due to measurement error. Taken together, these results support hypotheses 3 and 4.

Measure	Model	Mean Intercept	Mean Slope	Intercept Variance	Slope Variance	Standardised (Mean) Intercept-Slope Covariance
Between Errors	Basis	19.484***	-3.610***	131.539***	10.531***	-.735***
Delay Aversion	Basis	.244***	-.038***	.011***	.003***	-.665***
Deliberation Time	Basis	2.007***	-.237***	.216***	.031***	-.767***
Overall Proportion Bet	Linear	.491***	.011***	.011***	.002***	-.527***
Quality of Decision-Making	Basis	.937***	.013***	.004***	.001***	-.698***
Risk Adjustment	Basis	1.522***	.215***	.580***	.082***	-.304*
Risk Taking	Linear	.539***	.012***	.013***	.002***	-.528***
Strategy	Basis	31.286***	-1.749***	16.147***	2.304**	-.050
Surface Area	Basis	2.718***	.004	.088***	.005***	-.389***
Thickness	Basis	2.812***	-.072***	.016***	.004***	-.695***
Volume	Basis	8.923***	-.350***	1.107***	.153***	-.594***

Table 6. Mean, variance, and covariance estimates for intercepts and slopes of the winning (linear vs basis) behavioural and brain structure (grey matter) univariate models. * $p < .05$, ** $p < .01$, *** $p < .001$.

Bivariate models corroborate univariate findings for mean and variance estimates of intercepts and slopes

Next, I estimated bivariate latent growth curve models to quantify the associations between each behavioural and neural measure across time points (ages 14, 19, and 22 years). Similar to the univariate models, bivariate models had acceptable to good fit (Table 7), although RMSEA estimates and ranges were consistently worse (but still acceptable) in

models that included Surface Area. These results extend support for hypotheses 1 and 2 to include bivariate models. Note that, although the results for hypotheses 3 and 4 are shown for the bivariate models (Table 8; columns four and seven of Table 9), I do not discuss them as their values represent the same estimates (with minor discrepancies in magnitude) as the univariate models. An exception to this was Risk Taking-Volume, which had a non-significant brain slope variance (estimate = .045, $p > .05$), which was significant in the univariate model (Volume slope variance estimate = .153, $p < .001$, see column six of Table 6).

Behavioural Measure	**Brain Metric**	χ^2	**df**	**CFI**	**RMSEA**	**SRMR**	**AIC**	**BIC**
Between Errors	Surface Area***	74.84	7	.987	.065 [.052 .078]	.014	51271.855	51385.348
	Thickness*	18.58	7	.992	.028 [.013 .044]	.018	46321.192	46434.685
	Volume***	39.88	7	.989	.047 [.034 .062]	.020	42983.378	43096.880
Delay Aversion	Surface Area***	73.61	7	.984	.067 [.053 .081]	.018	-7445.307	-7331.870
	Thickness***	26.60	7	.982	.036 [.022 .052]	.022	-12225.075	-12111.648
	Volume***	44.99	7	.985	.050 [.037 .065]	.024	5108.825	5222.243
Deliberation Time	Surface Area***	102.87	7	.983	.071 [.059 .084]	.020	2940.499	3053.936
	Thickness***	36.91	7	.979	.044 [.031 .059]	.025	-1858.996	-1745.559
	Volume***	63.82	7	.980	.061 [.048 .075]	.027	15505.563	15618.981
Overall Proportion Bet	Surface Area***	135.40	8	.977	.080 [.068 .092]	.026	-8934.464	-8826.690
	Thickness***	54.80	8	.970	.053 [.040 .066]	.028	-13708.310	-13600.545
	Volume***	68.07	8	.979	.060 [.047 .073]	.029	3642.407	3750.163
Quality of Decision-making	Surface Area***	88.77	7	.985	.069 [.057 .083]	.014	-14651.406	-14537.960
	Thickness*	18.52	7	.993	.027 [.012 .042]	.017	-19444.322	-19330.886
	Volume***	34.74	7	.990	.045 [.031 .061]	.021	-2090.874	-1977.446
Risk Adjustment	Surface Area***	84.12	7	.986	.067 [.055 .080]	.017	9136.726	9250.171
	Thickness**	21.41	7	.992	.031 [.016 .046]	.017	4375.626	4489.062

	Volume***	38.41	7	.990	.045 [.032 .060]	.021	21688.028	21801.455
Risk Taking	Surface Area***	137.73	8	.977	.080 [.069 .092]	.026	-8478.744	-8370.970
	Thickness***	57.88	8	.969	.055 [.042 .068]	.029	-13253.139	-13145.374
	Volume***	40.61	7	.989	.048 [.034 .062]	.024	4070.185	4183.612
Strategy	Surface Area***	88.26	7	.986	.067 [.055 .080]	.014	24464.161	24577.682
	Thickness**	22.36	7	.991	.031 [.017 .047]	.020	40416.631	40530.142
	Volume***	41.60	7	.989	.046 [.033 .061]	.020	37045.282	37158.803

Table 7. Goodness-of-fit indices for bivariate models of behavioural measures and structural grey matter metrics. χ^2 = Chi-squared, df = degrees of freedom, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardised Root Mean Residual, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, * $p < .05$, ** $p < .01$, *** $p < .001$.

Bivariate Association		**Mean Intercept**		**Mean Slope**		**Intercept Variance**		**Slope Variance**	
Behavioural Measure	Brain Metric	Behaviour	Brain	Behaviour	Brain	Behaviour	Brain	Behaviour	Brain
Between Errors	Surface Area	19.486***	27.183***	-3.609***	.035	132.112***	8.835***	10.606***	.463***
	Thickness	19.492***	28.123***	-3.617***	-.719***	131.534***	1.595***	10.565***	.364***
	Volume	19.493***	8.922***	-3.616***	-.350***	132.046***	1.107***	10.586***	.154***
Delay Aversion	Surface Area	.243***	2.718***	-.038***	.004	.011***	.088***	.003***	.005***
	Thickness	.243***	2.812***	-.038***	-.072***	.012***	.016***	.003***	.004***
	Volume	.243***	8.924***	-.037***	-.350***	.011***	1.106***	.003***	.153***
Deliberation Time	Surface Area	2.007***	2.718***	-.236***	.004	.216***	.088***	.031***	.005***
	Thickness	2.008***	2.812***	-.237***	-.072***	.216***	.016***	.031***	.004***
	Volume	2.007***	8.923***	-.237***	-.350***	.216***	1.107***	.031***	.153***
Overall Proportion Bet	Surface Area	.492***	2.718***	.011***	.004	.011***	.089***	.002***	.005***
	Thickness	.491***	2.812***	.011***	-.072***	.011***	.016***	.002***	.004***
	Volume	.492***	8.923***	.011***	-.350***	.011***	1.106***	.002***	.154***

Quality of Decision-making	Surface Area	.937***	2.718***	.013***	.003	.004***	.088***	.001***	.005**
	Thickness	.937***	2.812***	.013***	-.072***	.004***	.016***	.001***	.004***
	Volume	.937***	8.923***	.013***	-.351***	.004***	1.107***	.001***	.154***
Risk Adjustment	Surface Area	1.524***	2.718***	.213***	.004	.572***	.088***	.078***	.005***
	Thickness	1.522***	2.812***	.214***	-.072***	.577***	.016***	.082***	.004***
	Volume	1.525***	8.924***	.212***	-.351***	.574***	1.107***	.079***	.153***
Risk Taking	Surface Area	.539***	2.718***	.012***	.004	.013***	.089***	.002***	.005***
	Thickness	.538***	2.812***	.013***	-.072***	.013***	.016***	.002***	.004***
	Volume	.539***	8.922***	.012***	-.356***	.013***	.721***	.002***	0.045
Strategy	Surface Area	31.284***	2.718***	-1.745***	.003	16.120***	.088***	2.267**	.005***
	Thickness	31.289***	28.123***	-1.748***	-.722***	16.056***	1.595***	2.280**	.363***
	Volume	31.285***	8.923***	-1.746***	-.350***	16.144***	1.107***	2.273**	.153***

Table 8. Mean and variance estimates for intercepts and slopes of the bivariate models.
* p < .05, ** p < .01, *** p < .001.

Bivariate latent growth models fail to support the structural scaffolding hypothesis

Next, I examined the covariances between the slopes and intercepts in the bivariate models. First, I observed that a subset of intercepts between behavioural abilities and brain structure were significant ($p < .05$, see column three of Table 9). Exceptions included Deliberation Time (no statistically significant covariances), as well as Between Errors and Risk Taking (all covariances with brain structure were statistically significant). There were several brain-behaviour intercept associations that were negative, implying a negative (although small) relationship between starting brain structure and behavioural performance for some measures: Between Errors-Volume, Delay Aversion-Surface Area, Delay Aversion-Volume, Risk Taking-Thickness, Strategy-Surface Area, and Strategy-Volume. For surface area and volume, these findings suggest that participants with larger baseline brain structure (i.e., greater volume and surface) also performed better on measures of decision-making and spatial working memory. Moreover, for cortical thickness, this interpretation suggests that

individuals with a thicker cortex at the onset of assessment were also better able to assess risk, again at age 14.

To statistically test for structural scaffolding of the cognitive measures (hypothesis 5), I inspected the covariance estimates for brain structure intercept and change/slope in behaviour. Doing this, I found no statistical evidence for structural scaffolding (Table 9, column six). In other words, contrary to my hypothesis, in this sample it was not the case that individual differences in brain structure at age 14 was associated with the rate of behavioural development across any measure of the three psychometric domains.

Next, I examined the reverse association: Does the current state of behavioural abilities (intercept) predict the rate of brain change (which was previously coined 'adaptive reorganisation', Kievit and Simpson-Kent, 2021). I observed several associations compatible with this hypothesis (see column five of Table 9 and Figure 22): For multiple bivariate models the intercept of the behavioural measure significantly predicted change in brain structure. This was found for three behaviour-brain couplings: Between Errors-Thickness (standardised covariance = .068, $p < .05$), Between Errors-Volume (standardised covariance = .116, $p < .01$), and Strategy-Volume (standardised covariance = .130, $p < .001$). This pattern suggests that, for these bivariate associations, baseline behavioural performance significantly predicted change in brain structure (i.e., cortical Thickness, and Volume) development. Moreover, the direction of these associations implies that, on average, lower starting scores (i.e., better performance) on spatial working memory (Between Errors and Strategy) led to a steeper (negative) brain change (e.g., more rapid thinning and/or larger volume loss in brain structure across waves). Finally, I did not observe correlated change (i.e., statistically significant covariances between behavioural and brain structure slopes). Therefore, in this sample and modelling framework, the rate of change at one level (e.g., behaviour) is not significantly associated with the rate of change in the other (e.g., brain).

Behavioural Measure	**Brain Metric**	**Covariance: Behaviour Intercept-Brain Intercept**	**Covariance: Behaviour Intercept-Behaviour Slope**	**Covariance: Behaviour Intercept-Brain Slope**	**Covariance: Brain Intercept-Behaviour Slope**	**Covariance: Brain Intercept-Brain Slope**	**Covariance: Behaviour Slope-Brain Slope**
Between Errors	Surface Area	.184***	-.733***	.066	-.075	-.388***	-.037

	Thickness	.073**	-.734***	.068*	-.081	-.696***	-.034
	Volume	-.143***	-.732***	.116**	.013	-.594***	-.059
Delay Aversion	Surface Area	-.082*	-.662***	.013	-.049	-.389***	.009
	Thickness	-.052	-.665***	.02	-.014	-.695***	.024
	Volume	-.085*	-.662***	.008	-.075	-.594***	.056
Deliberation Time	Surface Area	-.024	-.767***	.035	.008	-.389***	-.032
	Thickness	.016	-.767***	-.032	-.052	-.695***	.03
	Volume	-.018	-.767***	.013	-.008	-.594***	-.001
Overall Proportion Bet	Surface Area	.115***	-.528***	.036	.01	-.392***	-.055
	Thickness	-.056	-.525***	.003	.027	-.695***	-.042
	Volume	.096**	-.529***	.007	-.007	-.594***	-.051
Quality of Decision-making	Surface Area	.046	-.698***	-.044	.017	-.387***	.016
	Thickness	.047	-.698***	-.043	-.026	-.695***	.045
	Volume	.067*	-.697***	-.031	.003	-.594***	.001
Risk Adjustment	Surface Area	.154***	-.286*	-.042	.029	-.389***	-.009
	Thickness	.031	-.300*	-.03	.052	-.696***	-.059
	Volume	.164***	-.288*	-.059	.061	-.593***	-.071
Risk Taking	Surface Area	.124***	-.529***	.024	-.002	-.392***	-.037
	Thickness	-.064*	-.524***	.009	.039	-.695***	-.057
	Volume	.125**	-.530***	-.007	-.034	-.228	-.050
Strategy	Surface Area	-.241***	-.040	.077	-.034	-.389***	.011
	Thickness	.051	-.044	.046	-.054	-.695***	.043
	Volume	-.207***	-.040	.130***	-.057	-.594***	.019

Table 9. Standardised covariance estimates for bivariate models. * $p < .05$, ** $p < .01$, *** $p < .001$.

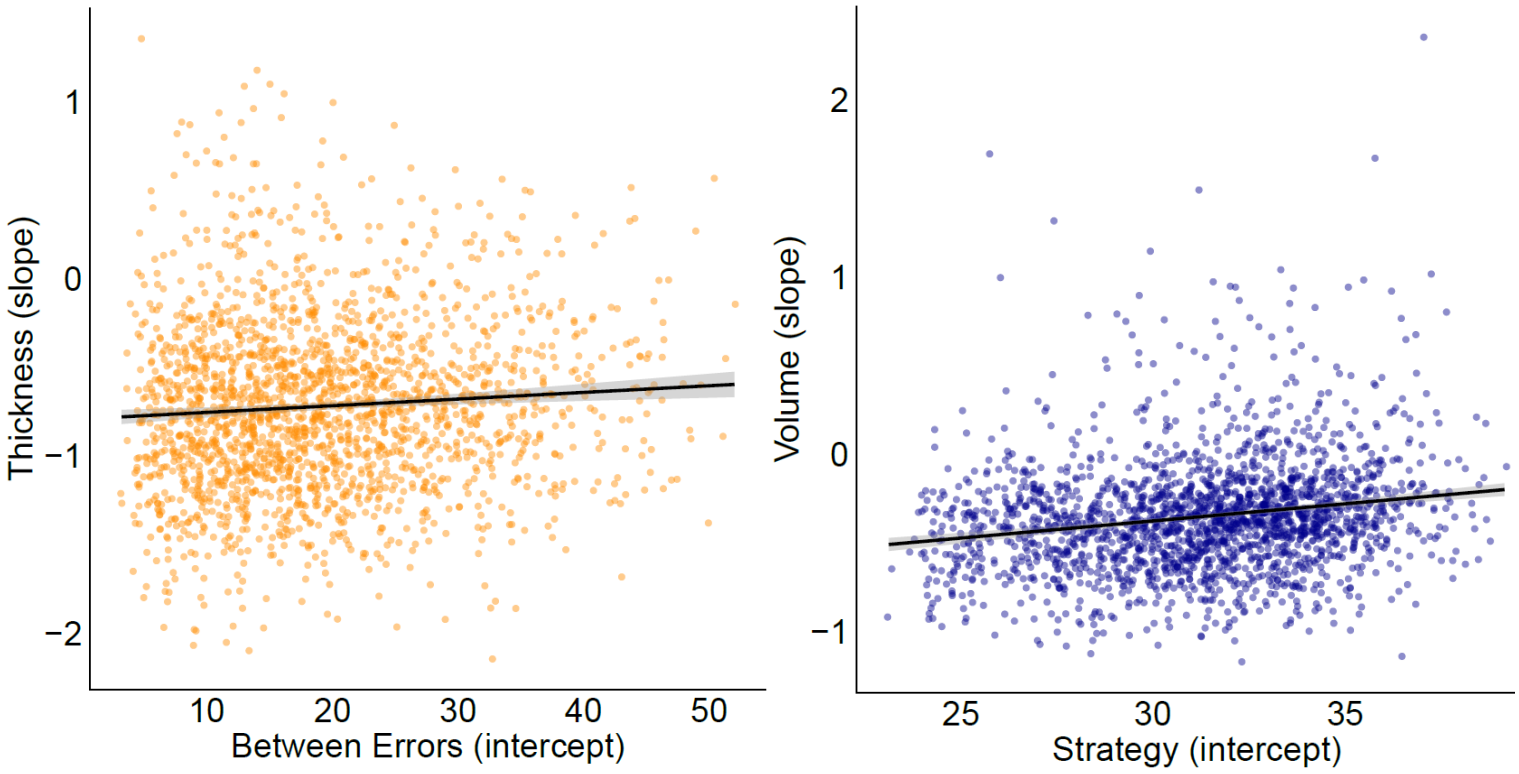


Figure 22. Left: Example of adaptive reorganisation using the predicted intercept and slope estimates of the Between Errors-Thickness model (i.e., lower intercept scores (better performance) on Between Errors predicted a faster rate of brain cortical thinning; r = .08). Right: Example of adaptive reorganisation using the predicted intercept and slope estimates of the Strategy-Volume model (i.e., lower intercept scores (better performance) on Strategy predicted a faster rate of brain cortical volume loss; r = .19).

Discussion

Summary and interpretation of findings

The present study sought to estimate the baseline, growth trajectories, and covariances between measures of decision-making, risk-related behaviours, and spatial working memory, as well as cortical brain structure (i.e., surface area, thickness, and volume) from mid-adolescence (age 14 years) to early adulthood (age 22 years). Using latent growth curve modelling, I found that most measures for behaviour (except for Overall Proportion Bet and Risk Taking) and all for neural structure change, and do so *nonlinearly*, across development (Figure 21). In addition, the univariate models also indicated that participants differed from one another both in baseline (age 14) and slope estimates (ages 14 – 22). This result held true for both the means (apart from Surface Area mean slope) and variances, implying that individuals not only started assessment at different behavioural and neural levels (i.e., baseline), but they also showed individual differences in patterns of change (i.e., slopes) across the time points included in this analysis (Table 6). When I inspected the mean intercept-slope covariances within domains, I also found that these relationships were universally significant (besides Strategy) and negatively correlated. This suggests that adolescents who began the IMAGEN study with lower behavioural performance and/or larger initial cortical brain structure tended to change more (i.e., behaviour: improved in performance, brain structure: showed less steep decrease) up until early adulthood. This is compatible with the growth models capturing differences in rates of maturation: Those who have already matured a lot (e.g., improved cognitive performance, or thinned their cortex) will, on average, change less rapidly, inducing a negative within-domain covariance.

Following these results, I fit bivariate latent growth curve models to estimate the correlations and covariances (i.e., intercepts and slopes) between each behavioural measure and brain structure metric. As expected, mean and variances estimates for intercepts and slopes were similar in magnitude and statistical significance as in the univariate models (Table 8). Finally, I examined brain-behaviour associations to statistically test for the presence of structural scaffolding, a phenomenon whereby baseline (intercept) cortical brain structure (whether surface area, thickness, and/or volume) predicts change (slope) in behavioural performance (e.g., spatial working memory). I found *no support* for structural scaffolding or correlated change (Table 9). Instead, I observed weak evidence in favour of adaptive

reorganisation, which states that the behavioural intercept(s) significantly predict change in brain structure in subsequent time points. This was found for three bivariate brain-behavioural models: Between Errors-Thickness, Between Errors-Volume, and Strategy-Volume (see Figure 22). This suggests that, on average, better starting scores on spatial working memory (i.e., Between Errors and Strategy) led to more rapid cortical thinning and/or volume loss across waves.

The observation that most behavioural (besides Overall Proportion Bet and Risk Taking) and all neural (i.e., cortical Surface Area, Thickness, and Volume) measures changed, and did so nonlinearly (i.e., increased behavioural performance and cortical shrinkage except for Surface Area) aligns with previous studies that demonstrated adolescence to be a period characterised by substantial alterations in behaviour (see Luna, 2009; Steinberg, 2005) and brain structure development (e.g., Becht and Mills, 2020; Tamnes et al., 2017). Therefore, strictly linear change (i.e., increase) and/or negative nonlinear change in behavioural abilities (except those similar to the ones shown to be linear in this study) and/or brain structure (e.g., cortical thinning) might serve as a potential neurocognitive marker of abnormal development, although one would need more waves of data to reliably predict this abnormality for a single individual. In other words, if a middle adolescent (e.g., ages 14 – 16 years) fails to show substantial behavioural development accompanied by abnormal (e.g., slow, or greater than average exponential) reductions in structural brain maturation, it could signify that they will face problems adjusting to future cognitive demands (e.g., success in undergraduate studies and/or occupational attainment).

The finding that Between Errors and Strategy, both spatial working memory measures, changed nonlinearly supports the notion that participants became significantly less error-prone and planned their choices better over successive assessments. This interpretation corroborates prior research, which demonstrated that working memory increases in childhood and adolescence (Best and Miller, 2010; Conklin et al., 2007; Huizinga et al., 2006). This coupled with the fact that Quality of Decision-making also showed pronounced nonlinear changes further hints that, in this sample, adolescents made *better* decisions and *improved* in cognitive ability up until early adulthood.

An alternative explanation for the nonlinear change in performance scores is a retest/practice effect (Scharfen et al., 2018). For instance, given that participants encountered

the same tasks across multiple waves, rather than nonlinear change in behavioural scores being the result of developmental factors (e.g., improved cognition and/or brain maturation), participants could have adjusted their strategy use (not to be confused with the Strategy measure used in this study) from mid-adolescence to early adulthood. On the other hand, instead of due to strategy adjustment, participants could have changed nonlinearly simply from being more familiar with the tasks across time points. Therefore, if change is truly linear, the presence of a retest/practice effect (whether through strategy adjustment and/or familiarity with the tasks) will make the trajectory appear nonlinear. While retest/practice effects were not estimated in this study, it is doubtful that strategy use and/or practice effects can account for the pattern of change in behavioural scores since the intervals between waves are no less than two years apart across the duration of the IMAGEN protocol (see https://imagen-europe.com/about/project/). Contrasting this point, a recent meta-analysis suggests that longer test-retest intervals (around 8 years) might be necessary to eliminate the influence of test-retest interval on retest/practice effects on cognitive tests (Scharfen et al., 2018). Therefore, future longitudinal studies should directly estimate the possible role of retest/practice effects before making strong claims about developmental change.

A final point worth discussing about nonlinear change in behavioural/cognitive performance is whether the degree of difficulty related to percentage changes in task performance across waves/time points is the 'same' at all ability levels. For example, is it easier to improve 5% in a task/measure when you previously scored below average (e.g., going from 50% to 55%) than it is to gain 5% points when your baseline performance is already near ceiling (e.g., increasing from 90% to 95%)? Although not tested in this study, an approach to answering this question is to use item response theory (IRT) modelling to establish an interval or ratio scale, which would enable one to confidently conclude whether behavioural development/change is truly nonlinear.

Overall Proportion Bet and Risk Taking showing linear change from ages 14 to 22 suggest that adolescents were relatively stable in their increase in risk-related behaviours. The only other risk-related behaviour, Risk Adjustment, although increasing nonlinearly, has an ambiguous interpretation (e.g., whether that change implies more erratic risk behaviour). Therefore, I conclude that from mid-adolescence to early adulthood, participants did not become worse at assessing risk. This interpretation agrees with recent perspectives that

challenge popular opinions and stereotypes of adolescents as purely irrational, instead providing a more nuanced view of adolescence as a formative transitional period into adulthood with both positive and negative consequences of risk-related behaviours (e.g., Blakemore, 2018; Do et al., 2020; Duell and Steinberg, 2019; Khurana et al., 2018; Maslowsky et al., 2019; Romer et al., 2017).

Interestingly, Surface Area, despite showing rapid growth from age 14 to 19 (slope was freely estimated), was the only variable with a non-significant mean intercept-slope covariance, which indicates that one *cannot* predict the trajectory of participants' cortical surface area from their baseline. This result hints that Surface Area has a distinct developmental pattern compared to Thickness and, hence, Volume (cortical volume is the product of surface area and thickness). Along these lines, recent research has suggested that the growth of cortical surface area and thickness (as well as and white matter maturation) is related to myelination. In particular, using functional, quantitative, and diffusion MRI as well as post-mortem histological methods, Natu et al., 2019 found that tissue growth in the ventral temporal cortex was associated with increased myelination from childhood to adulthood. This, in turn, affects the grey-white matter contrast in MRI scans, resulting in the apparent thinning of the cortex (e.g., see Becht and Mills, 2020; Tamnes et al., 2017; Vijayakumar et al., 2016). Therefore, and given that thickness and surface area were significantly (negatively) correlated in that study (i.e., Natu et al., 2019), increased myelination might lead to cortical thinning and a potential increase in surface area metrics.

In the bivariate behaviour-brain structure models, covariance estimates failed to support the structural scaffolding hypothesis. In other words, for all bivariate models, brain structure intercepts did not significantly predict changes in behavioural slopes across time points. There are at least three possible explanations for this finding. First, and perhaps the simplest, the structural scaffolding hypothesis may be false. Therefore, the pattern found in the studies (e.g., see Wendelken et al., 2017 and Ferrer, 2018) mentioned in Kievit and Simpson-Kent, 2021 might not generalise to other samples and cohorts. Second, in this study I used measures that are not typically associated with intelligence, except for spatial working memory (i.e., Between Errors and Strategy). Moreover, outside of spatial working memory, only Quality of Decision-making and Risk Adjustment have been found to positively associate with canonical intelligence measures, of which only Risk Adjustment significantly predicted

gains in IQ (Flouri et al., 2019). Moreover, in that study, Overall Proportion Bet was excluded from the analyses since it very strongly correlated with Risk Taking (r = .962). When I correlated these variables in the present dataset, I also found that Overall Proportion Bet and Risk Taking displayed a strong association (all r > .980 across waves), while no other abilities correlated with each other to that extent. In other words, the failure to find the predicted associations may be because of the particular measures used—I discuss this possibility in more detail below. Third and related to the second point, the 10 ROIs that I used to calculate the mean bilateral cortical brain structure were all regions known to be associated with traditional tests of intelligence. As a result, it might be possible that structural scaffolding does exists for these measures but requires a broader or alternative network of brain regions to discover its effect.

I observed some evidence for *adaptive reorganisation*, which predicts that behavioural intercept(s) will significantly covary with change in brain structure across waves. This effect was only found for three bivariate brain-behavioural models: Between Errors-Thickness, Between Errors-Volume, and Strategy-Volume. Furthermore, all three of these associations were positive. Therefore, given that higher values for these two abilities indicate worse performance, this relationship suggests that, on average, participants with better spatial working memory performance at baseline had more cortical volume loss and/or thinning from age 14 to 22. Adaptive reorganisation is in line with a study by Shaw et al., 2006, who found that individuals with higher IQ showed dynamic changes in cortical thickness from childhood to late adolescence. Specially, those with 'superior' intelligence underwent more pronounced cortical thickening in childhood, followed by intense cortical *thinning* by early adolescence (age 11.2 years). However, this study (IMAGEN) began assessment at age 14, after this 11.2-year-old cut-off. Therefore, even though the increased thinning continued until around age 16 (Shaw et al., 2006), further research is needed to substantiate adaptive reorganisation in adolescence. Finally, and surprisingly, there was no evidence for correlated change (significant covariances between behavioural and neural slopes). In close, in this cohort, behavioural performance and cortical brain structure did not significantly predict each other from mid-adolescence to early adulthood.

Limitations of the present study

Perhaps the most obvious limitation of the current study are the behavioural measures used. Structural scaffolding, adaptive reorganisation and correlated change, while possible to generalise to all behaviour-brain structure associations, were formulated as hypotheses about intelligence (Kievit and Simpson-Kent, 2021). Therefore, if these phenomena exist, their effect(s) should be most discoverable using assessments that have been verified to robustly correlate with IQ (e.g., fluid intelligence tasks such as Raven's Matrices, see Bilker et al., 2012). Therefore, future studies should examine these associations using more widely used tasks (e.g., working memory tasks such as digit recall, see Alloway, 2007). Furthermore, observation of these longitudinal bivariate hypotheses might require whole-brain (or more regionally specific) structure estimates rather than the reduced 10 ROIs used here (however, I do not test this here since whole-brain analyses were not part of the original conceptualisation and planning of the study). Relatedly, this study only incorporated grey matter metrics, although it is also known that white matter (e.g., fractional anisotropy, see Wandell, 2016) also significantly predict change in cognition from childhood to early adulthood (e.g., Wendelken et al., 2017).

Although the IMAGEN study has a substantially larger sample size compared to most neuroimaging studies (Poldrack et al., 2017), this cohort still contains fewer participants than might be necessary to find replicable results of brain-behaviour associations of small magnitude (Marek et al., 2020). As a result, future studies should seek to investigate these hypotheses in larger cohorts (e.g., ABCD, see Casey et al., 2018; Volkow et al., 2018) that surpass the "consortium" level for imaging ($N > 2000$, see Marek et al., 2020). Lastly, the time intervals between behavioural assessment and neuroimaging (i.e., $t2 - t1 = 5$ years; $t3 - t2 = 3$ years) might prevent the detection of neurocognitive changes that occur over a shorter time period (e.g., 1 year). Therefore, future studies should examine developmental changes in samples with more frequent behavioural and/or neural testing to help determine whether the patterns observed in this study using these measures translate to smaller developmental windows (e.g., each year between schooling).

Summary of the Chapter

In this chapter, I extended upon the findings from Chapter Two and Chapter Three by using longitudinal statistical methodology (i.e., latent growth curve modelling) to test *within-person* hypotheses of neurocognitive development (i.e., structural scaffolding, but also adaptive reorganisation and correlated change) from mid-adolescence to early adulthood (ages 14 – 22 years, using the IMAGEN study). I found no evidence for structural scaffolding or correlated change but did observe some support for *adaptive reorganisation* (current behavioural performance predicts change in brain structure). However, this result was only found for spatial working memory. Therefore, this finding suggests that, on average, *greater* performance on spatial working memory tasks (here measured by Between Errors and Strategy) led to *more rapid* cortical thinning and/or volume loss over time.

This concludes the empirical part of this dissertation. Now, in the final chapter, I summarise the main findings across the studies included in this thesis, address key limitations, and speculate about future directions towards advancing understanding of the developmental cognitive neuroscience of human intelligence.

Chapter Five

Concluding Remarks:

Summary, Limitations, and Future Directions

Summary of Thesis Findings

This dissertation has attempted to further understand the associations between human brain structure (i.e., grey matter and white matter) and intelligence (e.g., decision-making and working memory), particularly how they develop from early childhood to early adulthood (ages 5 – 22 years). These investigations were based on neurocognitive theory (i.e., crystallised and fluid intelligence, mutualism and network neuroscience, and structural scaffolding, adaptive reorganisation, and correlated change) and assessed in large (N range: 337 – 2072) developmental cohorts (i.e., CALM, IMAGEN, and NKI-Rockland) using well-established statistical methods (i.e., structural equation modelling and network science). Here I summarise the findings from each empirical chapter.

Chapter Two examined the latent variable structure and white matter determinants (i.e., fractional anisotropy) of child and adolescent intelligence (measured via tasks of crystallised and fluid intelligence, and working memory) using two cross-sectional, developmental samples (CALM: N = 551 (N = 165 neuroimaging), age range: 5 –18 years; NKI-Rockland: N = 337 (N = 65 neuroimaging), age range: 6 – 18 years). After estimating and comparing a series of structural equation models (SEM, see Kline, 2015), I found in both samples that, rather than the *g* factor, cognitive ability is best modelled as two separable but highly correlated constructs, crystallised (*gc*) and fluid (*gf*) intelligence. Interestingly, the best-fitting model for the CALM sample occurred when the working memory latent variable (LV) was assigned as an indicator of fluid intelligence, while for NKI-Rockland, the model fit best when working memory and *gf* were combined into a single LV, which has been suggested in prior research when participants are under increased time constraints (Chuderski, 2015, 2013). Moreover, in line with the age differentiation hypothesis (de Mooij et al., 2018; Garrett, 1946; Hülür et al., 2011), the covariance between *gc* and *gf* decreased from childhood to adolescence. This result was revealed through measurement invariance testing across age groups (Putnick

and Bornstein, 2016), which revealed that *gc* and *gf* became more distinct (i.e., less correlated) as the age of participants increased (i.e., cross-sectional childhood scores to cross-sectional adolescent scores). In terms of brain structure, I used Multiple Indicator Multiple Cause (MIMIC) analysis, which indicated that individual differences in white matter fractional anisotropy, especially of the superior longitudinal fasciculus, was strongly associated with *gc* and *gf* abilities. Finally, using a novel analysis framework, SEM trees (Brandmaier et al., 2013), a method that combines traditional SEM with decision trees, provided evidence for *neurocognitive reorganisation* of *gc* and *gf* and their white matter substrates. Hence, the relationships between *gc*, *gf* and a subset of white matter tracts (i.e., anterior thalamic radiations, cingulate gyrus, forceps major, forceps minor, and the superior longitudinal fasciculus, see Table 3) decreased between ages 7 – 8 years before increasing around age 10. Together, these results suggested that shortly before puberty marks a pivotal phase of change in the neurocognitive architecture of intelligence.

Chapter Three complemented and extended the analyses done in Chapter Two and sought to assess the mutualism theory of cognitive ability (van der Maas et al., 2006). Mutualism is a network model of general intelligence (see van der Maas et al., 2021, 2017) that claims that the positive manifold and *g* factor (Spearman, 1927, 1904) arise from direct (statistical) interactions among cognitive abilities such as reasoning and vocabulary throughout development (Kievit et al., 2019, 2017). Instead of using the SEM framework, which presumes the existence of unobserved LVs, this project used techniques from network science (Barabási, 2016), specifically network psychometrics (Epskamp et al., 2018). This analytic framework conceptualises psychological constructs such as general intelligence as a complex system and uses partial correlations, which estimate statistical dependencies among variables of interest (see also Epskamp et al., 2018). Hence, Chapter Three simultaneously modelled the brain-behaviour relationships essential for general intelligence in the same cohort as in Chapter Two (i.e., the CALM sample, see Holmes et al., 2019) but with more participants (ages 5 – 18 years; behavioural, N = 805; cortical volume, N = 246), except for white matter (fractional anisotropy, N = 165), which was the same as Chapter Two. The results of this analysis revealed that both single-layer (cognitive or neural nodes) and multilayer (combined cognitive and neural variables) networks contained mainly *small and positive* partial correlations, providing further support for mutualism theory (Kan et al., 2019; Schmank et al.,

2019). Furthermore, I attempted to identify possible groups of cognitive and/or neural nodes. To do this, I used community detection, particularly the Walktrap algorithm (Pons and Latapy, 2005) and calculated node centrality, both absolute strength (Bringmann et al., 2019) and bridge strength (see Jones et al., 2019). I found convergent evidence that certain cognitive (e.g., Reading) and neural (e.g., superior frontal gyrus) nodes may have had greater influence among (or might have been more influenced by) other nodes within the neurocognitive network. Together, these findings suggest that specific behavioural and neural variables function as *intermediary* nodes 'between' brain structure and cognitive ability.

Lastly, Chapter Four sought to statistically test for evidence of the structural scaffolding hypothesis of intelligence (see Figures 7 and 20). To do this, I employed latent growth curve modelling (Duncan and Duncan, 2004) to estimate the longitudinal relationships between human cortical grey matter structure (i.e., mean bilateral cortical surface area, thickness, and volume), and measures of decision-making, risk-related behaviours and spatial working memory from adolescence to early adulthood in the IMAGEN study (maximum N across time points/waves = 2072; age range: 14 – 22 years). For univariate models (i.e., solely behavioural or neural measures), comparisons between a linear and basis model of developmental growth revealed that both behavioural performance (here decision-making, risk-related behaviours, and spatial working memory), and cortical brain structure changed *nonlinearly* from mid-adolescence (age 14) to early adulthood (age 22). Moreover, bivariate models, which combined behavioural and neural measures, displayed support for *adaptive reorganisation* (behavioural intercept/baseline predicts changes/slopes in brain structure). However, no evidence was observed for structural scaffolding (brain structure intercept/baseline predicts changes/slopes in behaviour). Lastly, findings also failed to support the phenomenon of correlated change, whereby rates of change (i.e., covariance between behavioural and neural slopes) significantly predict each other, between ages 14 and 22 years. Overall, results from this study further suggest that mid-adolescence to early adulthood marks a distinct developmental period of brain-behaviour changes related to spatial working memory and decision-making, and grey matter cortical structure.

Limitations of this Dissertation

First, besides Chapter Four (IMAGEN study), the rest of the empirical chapters (Chapters Two and Three) in this dissertation only analyse cross-sectional data (CALM and NKI-Rockland cohorts). However, that is not to say that cross-sectional data are uninformative. Rather, cross-sectional analyses must be *complemented* by longitudinal analyses to understand neurocognitive processes more fully (e.g., see Raz and Lindenberger, 2011 and Schmiedek et al., 2020). Therefore, while *cross-sectional* data is informative relating to *between-subject* differences (e.g., group-level differences in cognitive ability), *longitudinal* samples are required to help tease apart *within-subject* differences in neurocognition. As noted in Kievit and Simpson-Kent, 2021, studies with longitudinal data for both behavioural/cognitive measures and structural neural metrics (e.g., grey matter: cortical surface area, thickness and volume; and white matter: fractional anisotropy) from childhood to early adulthood are scarce, with many having been published within the last five years. This rise in studying longitudinal developmental cohorts suggests that this approach, while growing, is still in its infancy. Therefore, future research programs should build upon this trend for a truly developmental cognitive neuroscience of intelligence.

Second, the cohorts included in this dissertation (i.e., CALM, IMAGEN and NKI-Rockland), while relatively representative of childhood, adolescent and/or early adulthood persons for their respective territories (CALM: East and South East of England; IMAGEN: Western Europe; NKI-Rockland: United States), still consists entirely of participants from WEIRD (Western, Educated, Industrialised, Rich, and Democratic) countries, which are known to often not generalise to non-WEIRD populations (Henrich et al., 2010). Therefore, future research is needed to assess the replicability of these findings in traditionally under-represented groups (such as, but not limited to, the Global South and other traditionally under-sampled populations).

Third, cohorts also tend to be extremely broad in their age ranges (e.g., see list of cohorts mentioned in Walhovd et al., 2018). While ambitious in their scope, studies that examine long-term changes might miss more *fine-grained* developmental shifts. For instance, if adolescence, a period marked by distinct behavioural and neural development, is defined as occurring between the ages of 10 and 24 (Sawyer et al., 2018), it is perhaps more informative to further divide statistical and/or empirical investigation into shorter time frames (e.g., early

adolescence may show differences from mid- and late adolescence). Related to this point, many longitudinal designs assess change over periods of months or years, which might fail to capture very acute (e.g., minutes or days) changes in behaviour and/or brain structure. While collection of such frequent data of both behavioural and neural dynamics in large samples (e.g., N ≥ 2000, see Marek et al., 2020) is not feasible due to the extreme logistical, monetary, and labour-intensive nature of such an endeavour, future cohorts should aim to shorten the window of assessment to approximate the outcomes of drastic developmental alterations (interval between assessments < 6 months).

Last but not least, given the observational nature of this thesis (i.e., non-experimental), one cannot conclusively determine cause-effect relationships due to the presence of possible confounding variables (e.g., see Rohrer, 2018). For example, I cannot confidently claim that, on average, having better spatial working memory performance causes more rapid cortical thinning or volume loss (adaptive reorganisation, see Chapter Four). Instead, I would have to directly intervene on behaviour (e.g., through cognitive training) and then compare brain structure shortly before and after the intervention. Conversely, I could use brain stimulation combined with cognitive training to measure the presence or absence of transfer effects for various intelligence-related tasks (e.g., for *gf*, see Brem et al., 2018). Moreover, I could perform analyses on lesion participants to study the neural determinants of intelligence (e.g., how loss of specific brain regions affect *g*, see Barbey et al., 2014, 2012; Gläscher et al., 2010; Woolgar et al., 2018, 2010). Therefore, while large observational studies such as those found in this dissertation are suitable to discover general trends in brain-behaviour associations, they must be *complemented* with experimental, lesion-symptom mapping, and meta-analytic methodology to further test mechanistic hypotheses. Related to this point, there exist many additional behavioural domains (e.g., cognitive constructs related to intelligence such as creativity, see Jauk et al., 2013 and Karwowski et al., 2017) and neural measures (e.g., brain function via activity patterns and cell recordings, or other structural correlates obtained through lesion studies) that also can provide further insight into neurocognitive function. As a result, future studies should seek to incorporate related cognitive abilities, brain metrics, and experimental/meta-analytic designs to more comprehensively tease apart the causal links between cognition and brain mechanisms.

This is by no means an exhaustive list as any study, no matter how rigorous, will contain areas for improvement and motivations for further study. Rather, these four limitations act as guiding principles to attempt to overcome in future studies on the development of human brain structure and intelligence from childhood to early adulthood. To close, I will now discuss what I argue can help overcome these limitations in research on neurocognitive differences and mechanisms of intelligence: theory building.

Future Directions

Towards mechanistic theory building in developmental cognitive neuroscience

My results from Chapter Two and Three that suggest verbal abilities (e.g., Peabody Picture Vocabulary Test and Reading) rather than fluid intelligence might play a more pivotal role in the development of cognitive ability fit with the gradual progression in schooling. For example, before children can successfully be taught more advanced subjects (e.g., history, reading comprehension, etc.), they must first become competent in basic language faculties. In other words, it may be that verbal skills (e.g., reading and spelling) facilitate performance on abstract tests, even in the absence of direct knowledge-based task demands. Recent evidence has been found supporting this notion and suggest that verbal ability, particularly reading and vocabulary in relation to working memory and reasoning, might drive early cognitive development (Kievit et al., 2019; Peng et al., 2018; Zhang and Malatesha Joshi, 2020). Therefore, future studies could further examine whether greater verbal ability in early development facilitates greater acquisition of higher-level cognitive skills by lowering computational demands in working memory.

Moreover, the fact that the Numerical Operations task was also found to be central (tri-layer network only, see Chapter Three) should be expected since mathematics (e.g., arithmetic) also involves symbol manipulation, as does language. In terms of mutualism (van der Maas et al., 2017, 2006), future models (ideally in longitudinal samples) could test whether language and other symbolic abilities show progressively higher reciprocal associations during early development compared to other abilities until more complex cognitive abilities (i.e., fluid reasoning and working memory) develop in later childhood (also see Kievit et al., 2019 and Peng et al., 2018).

These insights help to build *coherent*, *rigorous*, and *testable* neurocognitive theories of intelligence that have lacked in the psychology and neuroscience literature (Fried, 2020; Levenstein et al., 2020). In doing so, cognitive neuroscientists can help mitigate the limitations mentioned above as they would have more tangible interpretations of their results. Therefore, I argue that future research should aim to incorporate data from different scales, not only temporal (e.g., development across the lifespan) but also levels of organisation (e.g., brain, behaviour, genetics, and the environment). Furthermore, results from different levels can more easily be interpreted if these datasets are analysed using a unified conceptual and

quantitative framework (e.g., complex systems theory, see Siegenfeld and Bar-Yam, 2020) that combines strengths from various statistical techniques. Last, and perhaps most important, cognitive neuroscientists must formulate mechanistic (e.g., Bertolero et al., 2018) and generative models (for instance, Akarca et al., 2020) to gain further insights from past and help guide future controlled experiments. Researchers must not shy away from but rather embrace the complexity of the brain and cognition (see Fried and Robinaugh, 2020 for similar argument for mental health research).

One such proposal mentioned in Chapter One that attempts to explain general intelligence using a complex systems approach is The Network Neuroscience Theory of Human Intelligence (NNTHI, Barbey, 2018). Barbey argues that general intelligence arises from the dynamic small-world typology of the brain, which permits transitions between 'regular' or 'easy-to-reach' network states (needed to access prior knowledge for specific abilities) and 'random' or 'difficult-to-reach' (required to integrate information for broad abilities) network states (i.e., as in network control theory, see Gu et al., 2015). Together, this constrained flexibility allows the brain to adapt to novel cognitive domains (e.g., in abstract reasoning) while still preserving access to previously learned skills (e.g., from schooling). While evidence supporting the NNTHI has been inconclusive so far (Girn et al., 2019), it is a step (also see Process Overlap Theory, Kovacs and Conway, 2016, also mentioned in Chapter One) in the right direction toward a *complex systems theory of human intelligence*. Intelligence is a complex system—to understand it, we must treat it as such.

Supplementary Material

The Supplementary Material of Chapter Two has been published (although appears in this thesis in a modified form to ensure consistent formatting and thesis coherence):
Simpson-Kent I. L., Fuhrmann D., Bathelt J., Achterberg J., Borgeest G. S., the CALM Team, Kievit R. A. (2020). Neurocognitive reorganization between crystallized intelligence, fluid intelligence and white matter microstructure in two age-heterogeneous developmental cohorts. *Developmental Cognitive Neuroscience, Special Issue: Flux 2018: Mechanisms of Learning & Plasticity*. doi: 10.1016/j.dcn.2019.100743

The Supplementary Material of Chapter Three is available as a preprint (although appears in this thesis in a modified form to ensure consistent formatting and thesis coherence):
Simpson-Kent I. L., Fried E.I., Akarca D., Mareva S., Bullmore E.T., the CALM Team, Kievit R. A. (2021). Bridging brain and cognition: A multilayer network analysis of brain structural covariance and general intelligence in a developmental sample of struggling learners. Accepted, *Journal of Intelligence*. Preprint: *Biorxiv*. doi: 10.1101/2020.11.15.383869
Note: Supplementary Table 3 was created for this dissertation but is not included in the preprint. Also, Supplementary Figures 5 and 6 are different from the preprint due to an error discovered after uploading.

Although first-person pronouns are used throughout the chapter, this work is the result of collaborative research projects. My contributions to the above publications are:

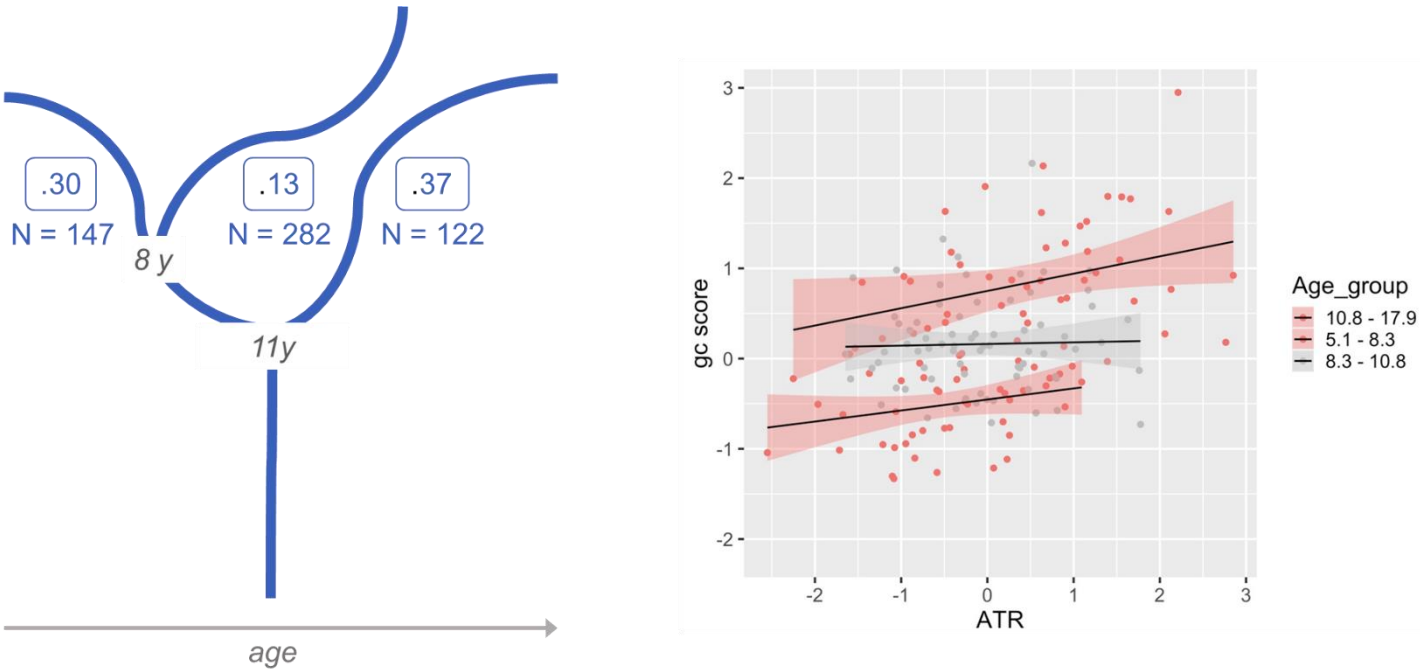
1. I led conceptualisation and planning (aided by R. A. Kievit and/or D. Fuhrmann) about the scientific hypotheses, analysis methods, and interpretations of the project.
2. I performed all manuscript analyses (aided by R. A. Kievit, D. Fuhrmann, Eiko I. Fried, and/or S. Mareva).
3. I wrote the first full draft (with input from R. A. Kievit and D. Fuhrmann) of the manuscript and led the revisions and confirmation of the final version (aided by other co-authors) of the manuscript for publication/preprint.

Supplementary Material for Chapter Two

Is the Peabody Picture Vocabulary Test a measure of fluid ability?

As a non-preregistered exploratory analysis, I more closely examined the cross-loading of the Peabody Picture Vocabulary Test (PPVT). This task asks participants to select the correct picture (out of four multiple-choice options) corresponding to the meaning of a word spoken by an examiner (Dunn and Dunn, 2007). As discussed previously in the **Results** section **Covariance among cognitive abilities cannot be captured by a single factor**, modification indices suggested the PPVT should either be cross-loaded or solely loaded onto *gf*. To better understand this cross-loading, I performed an exploratory (i.e., not part of preregistration, see https://aspredicted.org/5pz52.pdf) analysis using SEM tree analysis. In this analysis, I allowed the PPVT to load on both *gc* and *gf* and examined whether using age as a covariate yielded a developmental period where the associations between the latent factors and the PPVT task differed. This generated an age split for *gf* at around age 9.5 years whereby the loading of the PPVT decreased (from 1 to .87, unstandardised estimate).

Conversely, for *gc*, the loading remained the same (.12, unstandardised estimate). This suggested the PPVT as commonly implemented behaved as a fluid, rather than a crystallised, task, especially in younger participants of lower ability. A possible explanation for this pattern is that, while the PPVT draws on *gc*, the demanding nature of the task may require more fluid, executive components in younger children, especially in a cohort with comparatively low overall performance (e.g., CALM). Moreover, the surprisingly strong (.84, standardised) association between *gf* and PPVT in the full sample is similar to previous research in children (Naglieri, 1981) and adults (Bell et al., 2001), although with small, typically developing samples using different statistical methods.

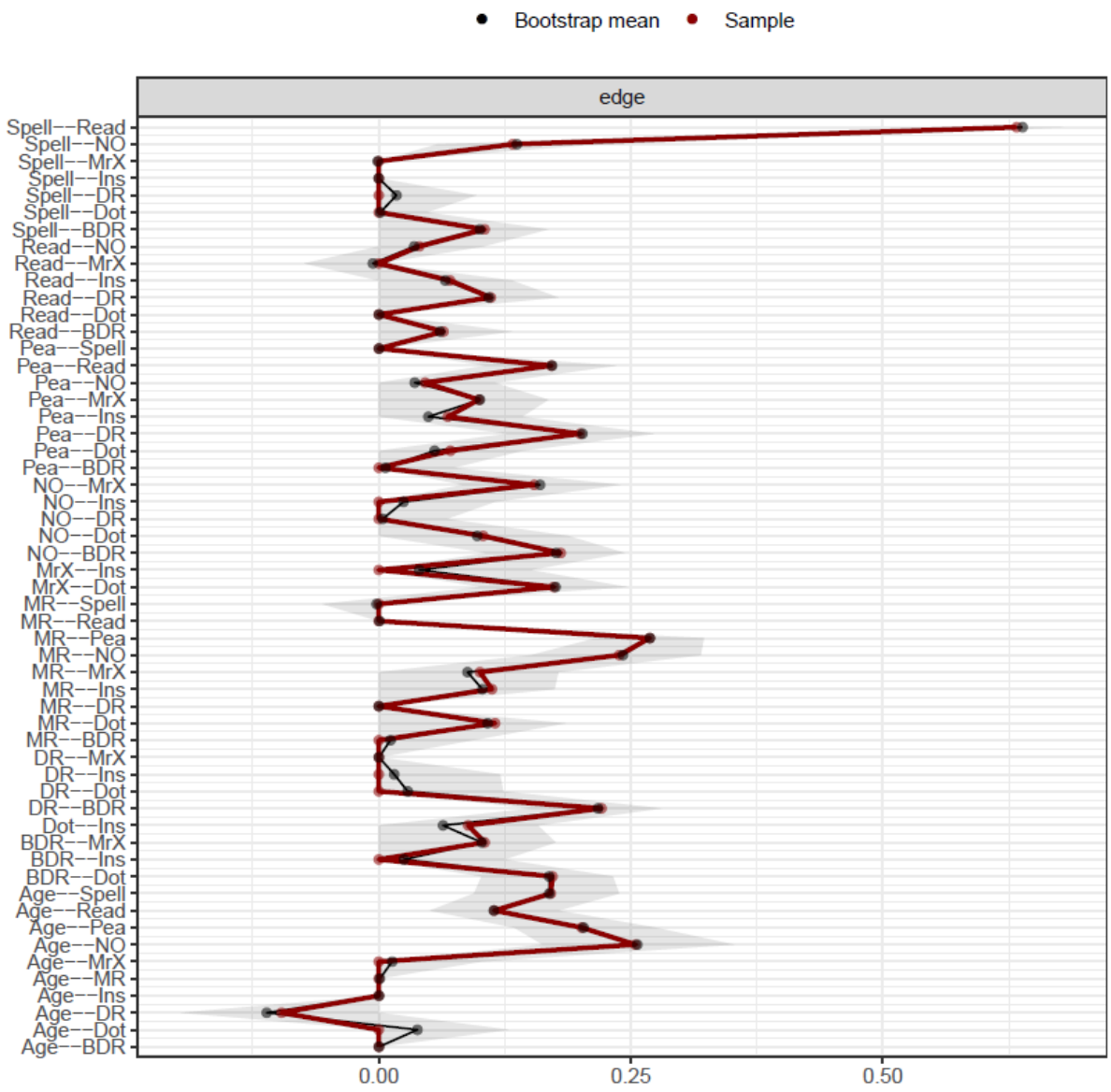


Supplementary Figure 1. Depiction of SEM tree analyses. Left: SEM tree results for the relationship between *gc* and the anterior thalamic radiations (ATR) (see Table 3 for full results). Figure adapted from Fuhrmann et al., 2020. Right: Visualisation of the nature of the effect for one path (ATR➔*gc*). The association between FA in the ATR and scores on the *gc* factor are moderately to strongly positive in the youngest children (r = .22) and the oldest children (r = .29), but effectively absent in the intermediate group (r = 0.02). Notably, the reader can see in this figure the steady increase in fractional anisotropy across ages (scaled ATR scores moving rightward) and improvement in *gc* (scaled *gc* scores moving upwards).

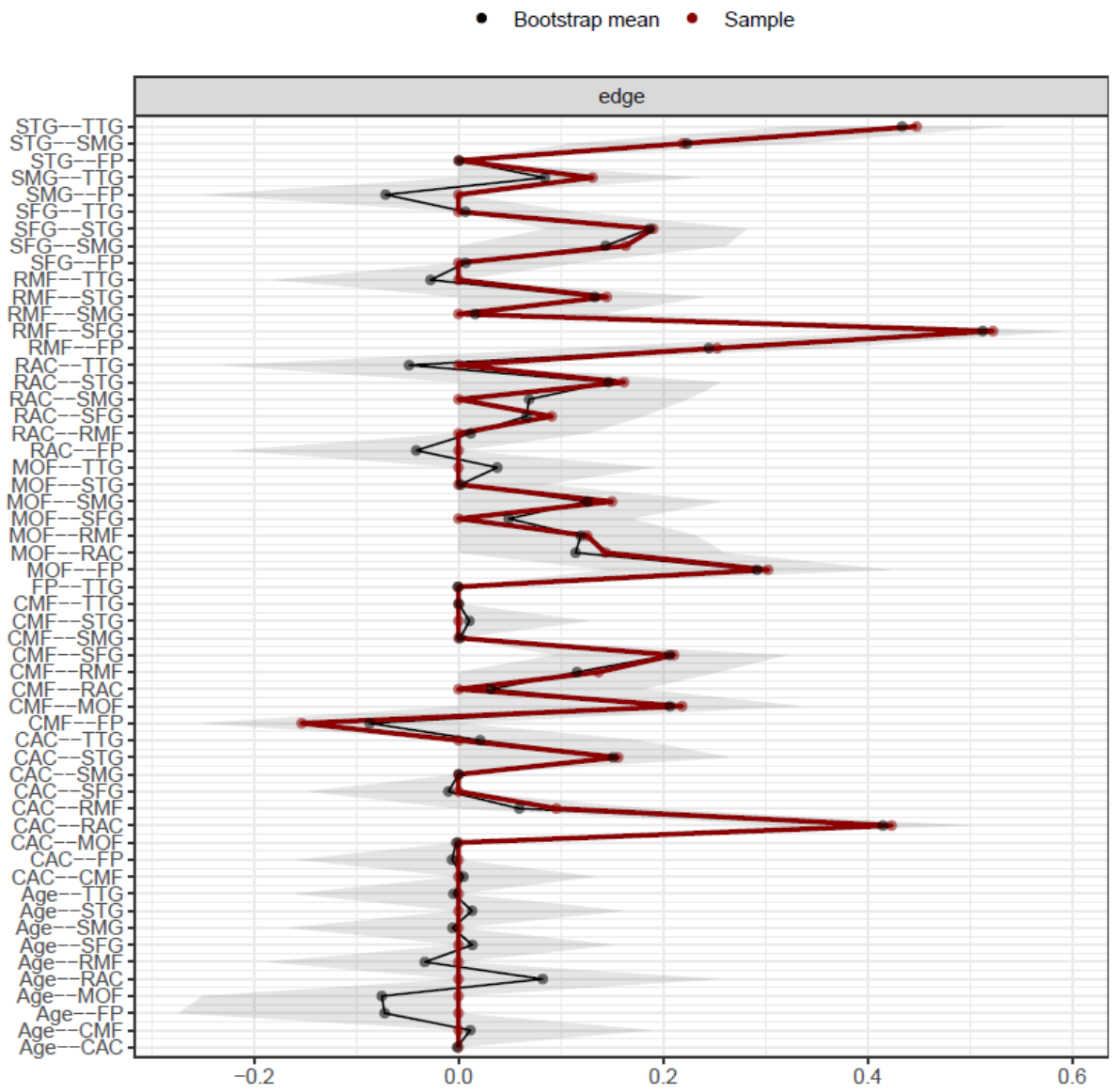
Supplementary Material for Chapter Three

Edge-weight stability analyses

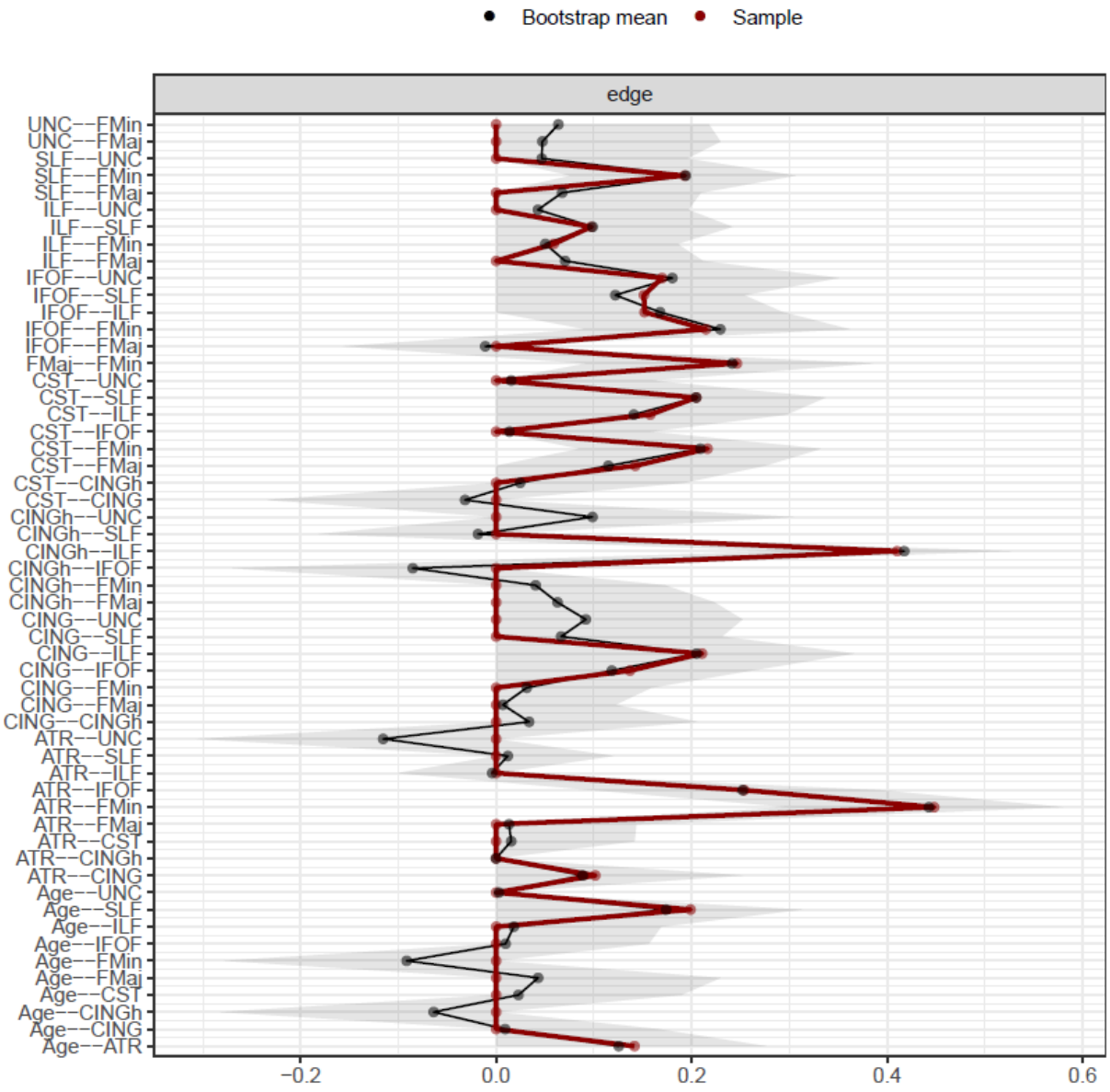
To further quantify the reliability of the partial correlation network edge-weights, I performed 2000 bootstraps and compared the bootstrapped mean values to the original sample estimates (Supplementary Figures 2 – 4). I do not show the bootstraps for the multilayer networks due to the size of the plots, but they can be found online (https://osf.io/36d2n/). Bootstrapped edge-weight means were consistently near the original sample value with the most variable being the white matter network (Supplementary Figure 4) and the multilayer networks (not shown). The low edge-weight stability in these networks could possibly be due to lower sample sizes of neural data (especially in the white matter network, N = 165, although centrality strength was moderately stable, CS-coefficient = 0.44), including when structural brain and cognitive data were combined. This, in turn, could have influenced the low stability estimates of the bridge centrality values in the multilayer networks.



Supplementary Figure 2. Comparisons between bootstrapped means and original sample edge-weight estimates for the CALM cognitive partial correlation network.



Supplementary Figure 3. Comparisons between bootstrapped means and original sample edge-weight estimates for the CALM grey matter partial correlation network.



Supplementary Figure 4. Comparisons between bootstrapped means and original sample edge-weight estimates for the CALM white matter partial correlation network.

The possible effect of outliers on major findings

In a previous version (https://www.biorxiv.org/content/10.1101/2020.11.15.383869v1) of this study, I observed that two FA values (one for the uncinate fasciculus, one for the forceps major), represented potential outliers with undue influence on the partitioning of the Walktrap algorithm in the single-layer white matter network. Removing this data yielded a distinct, and more parsimonious clustering solution (2 communities vs 5). Moreover, removing this outlier did not affect any summary statistics for the white matter partial correlation (single-layer) network except for range. Nevertheless, below I present the Pearson

correlations between the weights obtained from the original data presented in the main manuscript and those from the data after all outliers (defined as ± 4 standard deviations) are removed (Supplementary Table 1). Due to the vast similarity in descriptive statistics and high correlations between partial correlation weights, I conclude that outliers did not confound the results of this study. However, it must be noted that outliers might slightly affect community detection, but I chose to keep the original data due to the nature of the sample (struggling learners, therefore behavioural and neural data might be atypical to begin with) and given the fact that the neural data was already quality controlled. Furthermore, the two outlier white matter ROIs occurred in two separate participants (one outlier each) while the rest of their ROIs were consistent with the rest of the sample. In close, I argue that outliers (both cognitive and neural) are likely not due to measurement error but instead represent realistic values of an atypically developing sample.

Network Type	Original Data	Outliers Removed	Pearson Correlation
Cognitive	0.08 (0.11) [0, 0.63]	0.08 (0.11) [0, 0.61]	0.99
Grey Matter	0.09 (0.14) [-0.15, 0.52]	0.09 (0.14) [-0.15, 0.52]	1
White Matter	0.08 (0.11) [0, 0.44]	0.08 (0.13) [-0.14, 0.47]	0.93
Cognitive-grey matter	0.04 (0.1) [-0.12, 0.64]	0.03 (0.09) [-0.11, 0.62]	0.97
Cognitive-white matter	0.04 (0.1) [-0.2, 0.65]	0.04 (0.1) [-0.22, 0.65]	0.97
Tri-layer	0.02 (0.08) [-0.2, 0.66]	0.02 (0.08) [-0.19, 0.65]	0.98

Supplementary Table 1. Comparisons between partial correlation (PC) networks (original data vs outliers removed). These include summary statistics such as mean, (standard deviation), [range] and Pearson correlations between PC graph weights using pairwise complete observations to account for missingness.

How to deal with age?

As in previous literature, in the CALM sample age shows a clear positive association with intelligence measures (Figures 8 and 11) and brain structure (Figures 9 and 12). This fact, however, may further complicate any interpretations of (possible) causal interactions between cognitive and/or neural nodes. This is due to the multitude of reasons age might correlate with cognition and brain structure. For instance, this pattern could be due to the fact that older participants normally score higher on cognitive tasks and have greater brain maturation. In this case age functions as an underlying driver of (even greater) covariance between the two domains. There are at least two options (included in the original preprint) of how to deal with the relationship of age to cognitive ability, and grey and white matter structural covariance: 1) I could estimate the partial correlation network and include age as a node, therefore, choosing to estimate it *simultaneously* with the cognitive and neural variables (this is the option I chose for the non-Supplemental part of the analyses), or 2) I could *regress out* the association of age for each variable (age would show no correlation with cognitive and/or neural measures) *before* network estimation. Both approaches are related and have corresponding pros and cons. For example, these two options might enable the detection of correlations beyond age, possibility revealing *core relations* among variables independent of stereotypical neurocognitive development (e.g., older participants normally score higher on cognitive tasks and have larger brains as they mature). However, this might also remove developmental associations of interest (e.g., age may function as a moderator of cognitive and neural growth as in the above example).

Notably, a third possible option (assessed for this thesis and to be included in the revision of the preprint for future publication), which addresses this limitation, is to estimate the network *ignoring* age (i.e., *removing* it from dataset *before* estimation). Specifically, choosing not to include age as a node has the benefit of revealing the 'actual correlations' (i.e., those dependent on neurocognitive development in childhood and adolescence) among cognitive abilities and brain structure in the population, as the 'effects' of age are not controlled for before (regressed out) or during (age node associations with other nodes removed during calculation of partial correlations) network estimation. However, a drawback to this approach is that doing so could also amplify these associations, confounding the findings.

Here I compare the partial correlations matrices for the three analysis paths (i.e., age node used in network estimation vs age node regressed out before estimation; and age node used in network estimation vs age node removed from dataset prior to network estimation) for both single and multilayer networks (Supplementary Tables 2 and 3). This analysis demonstrates that, regardless of how age is accounted for in estimation, the partial correlation networks are very similar to each other.

Network Type	**Age Included in Estimation**	**Age Regressed Out before Estimation**	**Pearson Correlation**
Cognitive	0.08 (0.11) [0, 0.63]	0.08 (0.12) [0, 0.65]	0.98
Grey Matter	0.09 (0.14) [-0.15, 0.52]	0.09 (0.14) [-0.15, 0.52]	1(rounded from 0.999)
White Matter	0.08 (0.11) [0, 0.44]	0.08 (0.13) [-0.2, 0.49]	0.93
Cognitive-grey matter	0.04 (0.10) [-0.12, 0.64]	0.03 (0.10) [-0.14, 0.65]	0.94
Cognitive-white matter	0.04 (0.10) [-0.20, 0.65]	0.03 (0.10) [-0.24, 0.66]	0.94
Tri-layer	0.02 (0.08) [-0.20, 0.66]	0.02 (0.07) [0, 0.64]	0.88

Supplementary Table 2. Comparisons between partial correlation networks (age included in estimation vs age regressed out before estimation). These include summary statistics such as mean, (standard deviation), [range] and Pearson correlations between PC graph weights using pairwise complete observations to account for missingness.

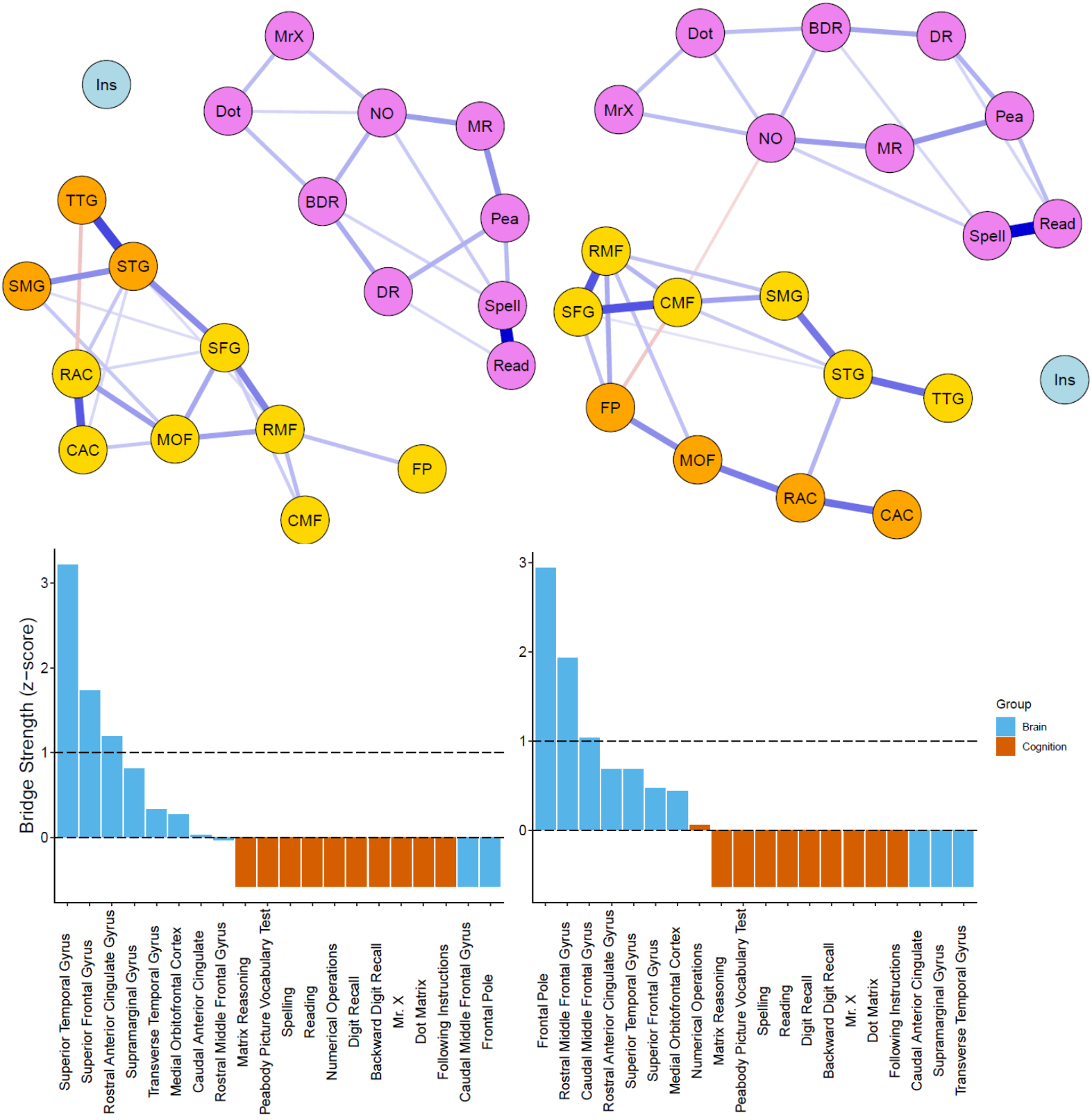
Network Type	Age Included in Estimation	Age Removed from Dataset before Estimation	Pearson Correlation
Cognitive	0.08 (0.11) [0, 0.63]	0.09 (0.12) [0, 0.68]	0.99
Grey Matter	0.09 (0.14) [-0.15, 0.52]	0.09 (0.14) [-0.16, 0.52]	0.99
White Matter	0.08 (0.11) [0, 0.44]	0.09 (0.13) [-0.19, 0.46]	0.90
Cognitive-grey matter	0.04 (0.10) [-0.12, 0.64]	0.04 (0.10) [-0.11, 0.66]	0.97
Cognitive-white matter	0.04 (0.10) [-0.20, 0.65]	0.04 (0.10) [-0.21, 0.69]	0.97
Tri-layer	0.02 (0.08) [-0.20, 0.66]	0.02 (0.08) [-0.16, 0.67]	0.94

Supplementary Table 3. Comparisons between partial correlation networks (age included in estimation vs age node removed from dataset prior to network estimation). These include summary statistics such as mean, (standard deviation), [range] and Pearson correlations between PC graph weights using pairwise complete observations to account for missingness.

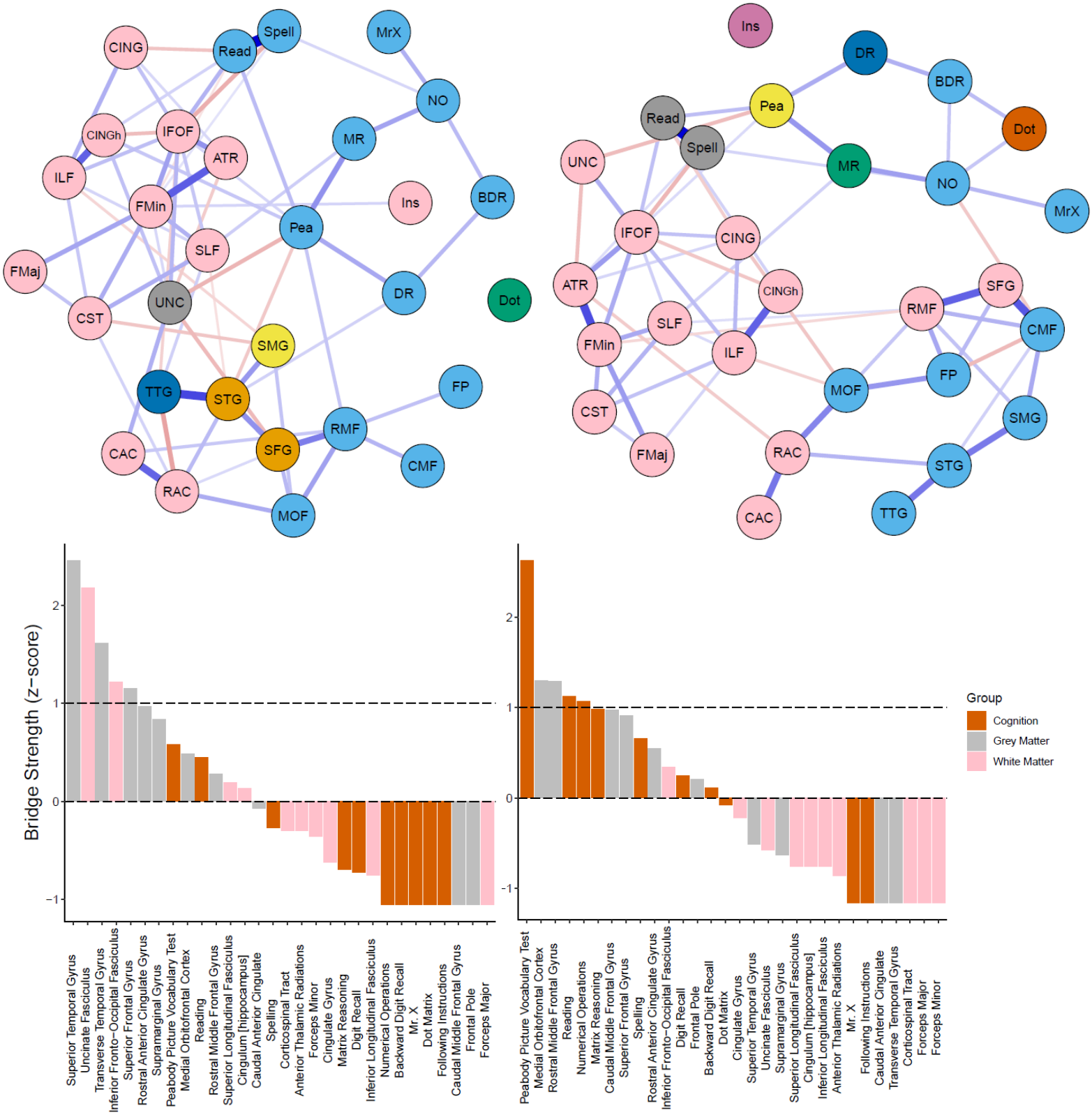
Teasing apart the relations of cortical volume to general intelligence:

Multilayer analysis using cortical surface area and thickness

Lastly, I partitioned cortical volume into its constituent parts, cortical surface area and thickness, to compare their partial correlations and community structures when combined with white matter and general intelligence (Supplementary Figures 5 and 6). This produced bilayer networks that were much less connected between domains (brain vs behaviour) than the cognition-volume bilayer network in Figure 15 (top left). Finally, bridge strength showed the same pattern as in the main manuscript, except for the surface area tri-layer network, where neural regions (both grey and white) appear to dominate the bridge strength centrality (Supplementary Figure 6), rather than cognition (Figure 16, bottom).



Supplementary Figure 5. Top: Network visualisations (spring layout) of partial correlation CALM bi-layer grey matter (surface area (left) and cortical thickness (right)) networks. Nodes are grouped according to Walktrap algorithm results. Bottom: Bridge centrality estimates (z-scores) for CALM bi-layer grey matter (surface area (left) and cortical thickness (right)) networks. Dashed lines indicate mean strength and one standard deviation above the mean.



Supplementary Figure 6. Top: Network visualisations (spring layout) of partial correlation CALM tri-layer grey matter (surface area (left) and cortical thickness (right)) networks. Nodes are grouped according to Walktrap algorithm results. Bottom: Bridge centrality estimates (z-scores) for CALM tri-layer grey matter (surface area (left) and cortical thickness (right)) networks. Dashed lines indicate mean strength and one standard deviation above the mean.

References

Aggarwal, I., Woolley, A.W., Chabris, C.F., Malone, T.W., 2015. Cognitive diversity, collective intelligence, and learning in teams.

Akaike, H., 1998. Information Theory and an Extension of the Maximum Likelihood Principle, in: Parzen, E., Tanabe, K., Kitagawa, G. (Eds.), Selected Papers of Hirotugu Akaike, Springer Series in Statistics. Springer, New York, NY, pp. 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15

Akarca, D., Vértes, P.E., Bullmore, E.T., Team, the C., Astle, D.E., 2020. A generative network model of neurodevelopment. bioRxiv 2020.08.13.249391. https://doi.org/10.1101/2020.08.13.249391

Alexander-Bloch, A., Giedd, J.N., Bullmore, E., 2013. Imaging structural co-variance between human brain regions. Nat Rev Neurosci 14, 322–336. https://doi.org/10.1038/nrn3465

Alloway, T.P., 2007. Automated Working Memory Assessment (AWMA).

Alloway, T.P., Gathercole, S.E., Kirkwood, H., Elliott, J., 2008. Evaluating the validity of the Automated Working Memory Assessment. Educational Psychology 28, 725–734. https://doi.org/10.1080/01443410802243828

Amestoy et al. See AUTHORS file (https://cran.r-project.org/web/packages/igraph/AUTHORS), 2020. igraph: Network Analysis and Visualization.

Andrews, J.L., Ahmed, S.P., Blakemore, S.-J., 2021. Navigating the Social Environment in Adolescence: The Role of Social Brain Development. Biological Psychiatry, Adolescent Brain Development and Psychopathology 89, 109–118. https://doi.org/10.1016/j.biopsych.2020.09.012

Arden, R., Adams, M.J., 2016. A general intelligence factor in dogs. Intelligence 55, 79–85. https://doi.org/10.1016/j.intell.2016.01.008

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis, Special Issue on The Third International Workshop on Biomedical Image Registration – WBIR 2006 12, 26–41. https://doi.org/10.1016/j.media.2007.06.004

Barabási, A.-L., 2016. Network Science. Cambridge University Press.

Barbey, A.K., 2018. Network Neuroscience Theory of Human Intelligence. Trends in Cognitive Sciences 22, 8–20. https://doi.org/10.1016/j.tics.2017.10.001

Barbey, A.K., Colom, R., Paul, E.J., Grafman, J., 2014. Architecture of fluid intelligence and working memory revealed by lesion mapping. Brain Struct Funct 219, 485–494. https://doi.org/10.1007/s00429-013-0512-z

Barbey, A.K., Colom, R., Solomon, J., Krueger, F., Forbes, C., Grafman, J., 2012. An integrative architecture for general intelligence and executive function revealed by lesion mapping. Brain 135, 1154–1164. https://doi.org/10.1093/brain/aws021

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I., 2001. The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. The Journal of Child Psychology

and Psychiatry and Allied Disciplines 42, 241–251. https://doi.org/10.1017/S0021963001006643
Bassett, D.S., Bullmore, E., 2006. Small-World Brain Networks. Neuroscientist 12, 512–523. https://doi.org/10.1177/1073858406293182
Bassett, D.S., Bullmore, E.T., 2017. Small-World Brain Networks Revisited. Neuroscientist 23, 499–516. https://doi.org/10.1177/1073858416667720
Bassett, D.S., Mattar, M.G., 2017. A Network Neuroscience of Human Learning: Potential to Inform Quantitative Theories of Brain and Behavior. Trends in Cognitive Sciences 21, 250–264. https://doi.org/10.1016/j.tics.2017.01.010
Bassett, D.S., Sporns, O., 2017. Network neuroscience. Nat Neurosci 20, 353–364. https://doi.org/10.1038/nn.4502
Bassett, D.S., Wymbs, N.F., Porter, M.A., Mucha, P.J., Carlson, J.M., Grafton, S.T., 2011. Dynamic reconfiguration of human brain networks during learning. PNAS 108, 7641–7646. https://doi.org/10.1073/pnas.1018985108
Bassett, D.S., Yang, M., Wymbs, N.F., Grafton, S.T., 2015. Learning-induced autonomy of sensorimotor systems. Nature Neuroscience 18, 744–751. https://doi.org/10.1038/nn.3993
Basten, U., Hilger, K., Fiebach, C.J., 2015. Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. Intelligence 51, 10–27. https://doi.org/10.1016/j.intell.2015.04.009
Bathelt, J., Geurts, H.M., Borsboom, D., 2020. More than the sum of its parts: Merging network psychometrics and network neuroscience with application in autism. bioRxiv 2020.11.17.386276. https://doi.org/10.1101/2020.11.17.386276
Bathelt, J., Johnson, A., Zhang, M., Astle, D.E., 2019. The cingulum as a marker of individual differences in neurocognitive development. Scientific Reports 9, 2281. https://doi.org/10.1038/s41598-019-38894-z
Beaujean, A.A., 2015. John Carroll's Views on Intelligence: Bi-Factor vs. Higher-Order Models. Journal of Intelligence 3, 121–136. https://doi.org/10.3390/jintelligence3040121
Becht, A.I., Mills, K.L., 2020. Modeling Individual Differences in Brain Development. Biological Psychiatry, Convergence and Heterogeneity in Psychopathology 88, 63–69. https://doi.org/10.1016/j.biopsych.2020.01.027
Bell, N.L., Lassiter, K.S., Matthews, T.D., Hutchinson, M.B., 2001. Comparison of the Peabody Picture Vocabulary Test—Third Edition and Wechsler Adult Intelligence Scale—Third Edition with university students. Journal of Clinical Psychology 57, 417–422. https://doi.org/10.1002/jclp.1024
Benson, N., 2008. Cattell—Horn—Carroll Cognitive Abilities and Reading Achievement. Journal of Psychoeducational Assessment 26, 27–41. https://doi.org/10.1177/0734282907301424
Bertolero, M.A., Yeo, B.T.T., Bassett, D.S., D'Esposito, M., 2018. A mechanistic model of connector hubs, modularity and cognition. Nature Human Behaviour 2, 765–777. https://doi.org/10.1038/s41562-018-0420-6
Bertolero, M.A., Yeo, B.T.T., D'Esposito, M., 2017. The diverse club. Nature Communications 8, 1277. https://doi.org/10.1038/s41467-017-01189-w
Bertolero, M.A., Yeo, B.T.T., D'Esposito, M., 2015. The modular and integrative functional architecture of the human brain. PNAS 112, E6798–E6807. https://doi.org/10.1073/pnas.1510619112

Best, J.R., Miller, P.H., 2010. A Developmental Perspective on Executive Function. Child Development 81, 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x
Betzel, R., 2020. Network neuroscience and the connectomics revolution. arXiv:2010.01591 [q-bio].
Betzel, R.F., Bassett, D.S., 2017. Multi-scale brain networks. NeuroImage, Functional Architecture of the Brain 160, 73–83. https://doi.org/10.1016/j.neuroimage.2016.11.006
Betzel, R.F., Gu, S., Medaglia, J.D., Pasqualetti, F., Bassett, D.S., 2016. Optimally controlling the human connectome: the role of network topology. Scientific Reports 6, 30770. https://doi.org/10.1038/srep30770
Bianconi, G., 2018. Multilayer Networks: Structure and Function. Oxford University Press.
Bickley, P.G., Keith, T.Z., Wolfle, L.M., 1995. The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. Intelligence 20, 309–328. https://doi.org/10.1016/0160-2896(95)90013-6
Bilker, W.B., Hansen, J.A., Brensinger, C.M., Richard, J., Gur, R.E., Gur, R.C., 2012. Development of Abbreviated Nine-item Forms of the Raven's Standard Progressive Matrices Test. Assessment 19, 354–369. https://doi.org/10.1177/1073191112446655
Blakemore, S.-J., 2018. Avoiding Social Risk in Adolescence. Curr Dir Psychol Sci 27, 116–122. https://doi.org/10.1177/0963721417738144
Blakemore, S.-J., 2008. The social brain in adolescence. Nat Rev Neurosci 9, 267–277. https://doi.org/10.1038/nrn2353
Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., Fox, J., 2011. OpenMx: An Open Source Extended Structural Equation Modeling Framework. Psychometrika 76, 306–317. https://doi.org/10.1007/s11336-010-9200-6
Borsboom, D., 2017. A network theory of mental disorders. World Psychiatry 16, 5–13. https://doi.org/10.1002/wps.20375
Borsboom, D., Mellenbergh, G., Heerden, J., 2003. The Theoretical Status of Latent Variables. Psychological review 110, 203–19. https://doi.org/10.1037/0033-295X.110.2.203
Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika 52, 345–370. https://doi.org/10.1007/BF02294361
Brandmaier, A.M., Prindle, J.J., McArdle, J.J., Lindenberger, U., 2016. Theory-guided exploration with structural equation model forests. Psychological Methods 21, 566–582. https://doi.org/10.1037/met0000090
Brandmaier, A.M., von Oertzen, T., McArdle, J.J., Lindenberger, U., 2013. Structural equation model trees. Psychological Methods 18, 71–86. https://doi.org/10.1037/a0030001
Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., Schweiger, J.I., Grimm, O., Heinz, A., Tost, H., Meyer-Lindenberg, A., Bassett, D.S., 2015. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. PNAS 112, 11678–11683. https://doi.org/10.1073/pnas.1422487112
Brem, A.-K., Almquist, J.N.-F., Mansfield, K., Plessow, F., Sella, F., Santarnecchi, E., Orhan, U., McKanna, J., Pavel, M., Mathan, S., Yeung, N., Pascual-Leone, A., Kadosh, R.C., Brem, A.-K., Kadosh, R.C., Mansfield, K., Yeung, N., Plessow, F., Santarnecchi, E., Pascual-Leone, A., Almquist, J.N.-F., Dillard, M., Orhan, U., Mathan, S., McKanna, J., Erdogmus, D., Pavel, M., Kimball, G., Myers, E., 2018. Modulating fluid intelligence performance through combined cognitive training and brain stimulation.

Neuropsychologia, The neural bases of creativity and intelligence: common grounds and differences 118, 107–114. https://doi.org/10.1016/j.neuropsychologia.2018.04.008
Bringmann, L.F., Elmer, T., Epskamp, S., Krause, R.W., Schoch, D., Wichers, M., Wigman, J.T.W., Snippe, E., 2019. What do centrality measures measure in psychological networks? Journal of Abnormal Psychology 128, 892–903. https://doi.org/10.1037/abn0000446
Bullmore, E., Sporns, O., 2012. The economy of brain network organization. Nature Reviews Neuroscience 13, 336–349. https://doi.org/10.1038/nrn3214
Burgoyne, A.P., Hambrick, D.Z., Altmann, E.M., 2019. Is working memory capacity a causal factor in fluid intelligence? Psychon Bull Rev. https://doi.org/10.3758/s13423-019-01606-9
Calvin, C.M., Deary, I.J., Fenton, C., Roberts, B.A., Der, G., Leckenby, N., Batty, G.D., 2011. Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. International Journal of Epidemiology 40, 626–644. https://doi.org/10.1093/ije/dyq190
Canivez, G.L., Youngstrom, E.A., 2019. Challenges to the Cattell-Horn-Carroll Theory: Empirical, Clinical, and Policy Implications. Applied Measurement in Education 32, 232–248. https://doi.org/10.1080/08957347.2019.1619562
Cannon, T.D., Keller, M.C., 2006. Endophenotypes in the Genetic Analyses of Mental Disorders. Annu. Rev. Clin. Psychol. 2, 267–290. https://doi.org/10.1146/annurev.clinpsy.2.022305.095232
Carroll, J.B., 1997. The three-stratum theory of cognitive abilities, in: Contemporary Intellectual Assessment: Theories, Tests, and Issues. The Guilford Press, New York, NY, US, pp. 122–130.
Carroll, J.B., 1993. Human Cognitive Abilities: A Survey of Factor-Analytic Studies. Cambridge University Press.
Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Chaarani, B., Mejia, M.H., Hagler, D.J., Daniela Cornejo, M., Sicat, C.S., Harms, M.P., Dosenbach, N.U.F., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J.R., Kuperman, J.M., Fair, D.A., Dale, A.M., 2018. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Developmental Cognitive Neuroscience, The Adolescent Brain Cognitive Development (ABCD) Consortium: Rationale, Aims, and Assessment Strategy 32, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001
Cattell, R.B., 1971. Abilities: their structure, growth, and action, Abilities: their structure, growth, and action. Houghton Mifflin, Oxford, England.
Cattell, R.B., 1967. The theory of fluid and crystallized general intelligence checked at the 5-6 year-old level. British Journal of Educational Psychology 37, 209–224. https://doi.org/10.1111/j.2044-8279.1967.tb01930.x
Cattell, R.B., 1963. Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology 54, 1–22. https://doi.org/10.1037/h0046743
Chuderski, A., 2015. The broad factor of working memory is virtually isomorphic to fluid intelligence tested under time pressure. Personality and Individual Differences 85, 98–104. https://doi.org/10.1016/j.paid.2015.04.046

Chuderski, A., 2013. When are fluid intelligence and working memory isomorphic and when are they not? Intelligence 41, 244–262. https://doi.org/10.1016/j.intell.2013.04.003
Cole, M.W., Reynolds, J.R., Power, J.D., Repovs, G., Anticevic, A., Braver, T.S., 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. Nature Neuroscience 16, 1348–1355. https://doi.org/10.1038/nn.3470
Conklin, H.M., Luciana, M., Hooper, C.J., Yarger, R.S., 2007. Working Memory Performance in Typically Developing Children and Adolescents: Behavioral Evidence of Protracted Frontal Lobe Development. Developmental Neuropsychology 31, 103–128. https://doi.org/10.1207/s15326942dn3101_6
Conway, A.R.A., Cowan, N., Bunting, M.F., Therriault, D.J., Minkoff, S.R.B., 2002. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. Intelligence 30, 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4
Conway, A.R.A., Kane, M.J., Engle, R.W., 2003. Working memory capacity and its relation to general intelligence. Trends in Cognitive Sciences 7, 547–552. https://doi.org/10.1016/j.tics.2003.10.005
Dablander, F., Hinne, M., 2019. Node centrality measures are a poor substitute for causal inference. Scientific Reports 9, 1–13. https://doi.org/10.1038/s41598-019-43033-9
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage 9, 179–194. https://doi.org/10.1006/nimg.1998.0395
Damerius, L.A., Burkart, J.M., van Noordwijk, M.A., Haun, D.B.M., Kosonen, Z.K., Galdikas, B.M.F., Saraswati, Y., Kurniawan, D., van Schaik, C.P., 2019. General cognitive abilities in orangutans (Pongo abelii and Pongo pygmaeus). Intelligence, Intelligence in non human animals 74, 3–11. https://doi.org/10.1016/j.intell.2018.10.008
De Domenico, M., Sasai, S., Arenas, A., 2016. Mapping Multiplex Hubs in Human Functional Brain Networks. Front. Neurosci. 10. https://doi.org/10.3389/fnins.2016.00326
de Mooij, S.M.M., Henson, R.N.A., Waldorp, L.J., Kievit, R.A., 2018. Age Differentiation within Gray Matter, White Matter, and between Memory and White Matter in an Adult Life Span Cohort. The Journal of Neuroscience 38, 5826–5836. https://doi.org/10.1523/JNEUROSCI.1627-17.2018
Deakin, J., Aitken, M., Robbins, T., Sahakian, B.J., 2004. Risk taking during decision-making in normal volunteers changes with age. Journal of the International Neuropsychological Society 10, 590–598. https://doi.org/10.1017/S1355617704104104
Deary, I.J., Penke, L., Johnson, W., 2010. The neuroscience of human intelligence differences. Nature Reviews Neuroscience 11, 201–211. https://doi.org/10.1038/nrn2793
Deary, I.J., Strand, S., Smith, P., Fernandes, C., 2007. Intelligence and educational achievement. Intelligence 35, 13–21. https://doi.org/10.1016/j.intell.2006.02.001
Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021
Diachek, E., Blank, I., Siegelman, M., Affourtit, J., Fedorenko, E., 2020. The Domain-General Multiple Demand (MD) Network Does Not Support Core Aspects of Language

Comprehension: A Large-Scale fMRI Investigation. J. Neurosci. 40, 4536–4550. https://doi.org/10.1523/JNEUROSCI.2036-19.2020
Dickens, W.T., Flynn, J.R., 2001. Heritability estimates versus large environmental effects: The IQ paradox resolved. Psychological Review 108, 346–369. https://doi.org/10.1037/0033-295X.108.2.346
Do, K.T., Sharp, P.B., Telzer, E.H., 2020. Modernizing Conceptions of Valuation and Cognitive-Control Deployment in Adolescent Risk Taking. Curr Dir Psychol Sci 29, 102–109. https://doi.org/10.1177/0963721419887361
Ducharme, S., Albaugh, M.D., Nguyen, T.-V., Hudziak, J.J., Mateos-Pérez, J.M., Labbe, A., Evans, A.C., Karama, S., 2015. Trajectories of cortical surface area and cortical volume maturation in normal brain development. Data in Brief 5, 929–938. https://doi.org/10.1016/j.dib.2015.10.044
Duell, N., Steinberg, L., 2019. Positive Risk Taking in Adolescence. Child Development Perspectives 13, 48–52. https://doi.org/10.1111/cdep.12310
Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends in Cognitive Sciences 14, 172–179. https://doi.org/10.1016/j.tics.2010.01.004
Duncan, J., Assem, M., Shashidhara, S., 2020. Integrated Intelligence from Distributed Brain Activity. Trends in Cognitive Sciences 24, 838–852. https://doi.org/10.1016/j.tics.2020.06.012
Duncan, T.E., Duncan, S.C., 2004. An introduction to latent growth curve modeling. Behavior Therapy 35, 333–363. https://doi.org/10.1016/S0005-7894(04)80042-X
Dunn, L.M., Dunn, D.M., 2007. PPVT-4: Peabody picture vocabulary test.
Dux, P.E., Ivanoff, J., Asplund, C.L., Marois, R., 2006. Isolation of a Central Bottleneck of Information Processing with Time-Resolved fMRI. Neuron 52, 1109–1120. https://doi.org/10.1016/j.neuron.2006.11.009
Eaton, D.K., Kann, L., Kinchen, S., Ross, J., Hawkins, J., Harris, W.A., Lowry, R., McManus, T., Chyen, D., Shanklin, S., Lim, C., Grunbaum, J.A., Wechsler, H., 2006. Youth Risk Behavior Surveillance—United States, 2005. Journal of School Health 76, 353–372. https://doi.org/10.1111/j.1746-1561.2006.00127.x
Edelman, G.M., Gally, J.A., 2001. Degeneracy and complexity in biological systems. PNAS 98, 13763–13768. https://doi.org/10.1073/pnas.231499798
Enders, C.K., 2001. The Performance of the Full Information Maximum Likelihood Estimator in Multiple Regression Models with Missing Data. Educational and Psychological Measurement 61, 713–740. https://doi.org/10.1177/0013164401615001
Engle, R.W., Tuholski, S.W., Laughlin, J.E., Conway, A.R.A., 1999. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. Journal of Experimental Psychology: General 128, 309–331. https://doi.org/10.1037/0096-3445.128.3.309
Epskamp, S., Borsboom, D., Fried, E.I., 2018. Estimating psychological networks and their accuracy: A tutorial paper. Behavior Research Methods 50, 195–212. https://doi.org/10.3758/s13428-017-0862-1
Epskamp, S., Costantini, G., Haslbeck, J., Isvoranu, A., Cramer, A.O.J., Waldorp, L.J., Schmittmann, V.D., Borsboom, D., 2020. qgraph: Graph Plotting Methods, Psychometric Data Visualization and Graphical Model Estimation.
Epskamp, S., Fried, E.I., 2020. bootnet: Bootstrap Methods for Various Network Estimation Routines.

Epskamp, S., Fried, E.I., 2018. A tutorial on regularized partial correlation networks. Psychological Methods 23, 617–634. https://doi.org/10.1037/met0000167
Fedorenko, E., Duncan, J., Kanwisher, N., 2013. Broad domain generality in focal regions of frontal and parietal cortex. PNAS. https://doi.org/10.1073/pnas.1315235110
Ferrer, E., 2018. Discrete-and semi-continuous time latent change score models of fluid reasoning development from childhood to adolescence, in: Longitudinal Multivariate Psychology. Routledge.
Ferrer, E., McArdle, J.J., 2004. An Experimental Analysis of Dynamic Hypotheses About Cognitive Abilities and Achievement From Childhood to Early Adulthood. Developmental Psychology 40, 935–952. https://doi.org/10.1037/0012-1649.40.6.935
Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. PNAS 97, 11050–11055. https://doi.org/10.1073/pnas.200033797
Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. Neuron 33, 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. NeuroImage 9, 195–207. https://doi.org/10.1006/nimg.1998.0396
Flanagan, D.P., 2000. Wechsler-based CHC cross-battery assessment and reading achievement: Strengthening the validity of interpretations drawn from Wechsler test scores. School Psychology Quarterly 15, 295–329. https://doi.org/10.1037/h0088789
Flanagan, D.P., Dixon, S.G., 2014. The Cattell-Horn-Carroll Theory of Cognitive Abilities, in: Encyclopedia of Special Education. American Cancer Society. https://doi.org/10.1002/9781118660584.ese0431
Flouri, E., Moulton, V., Ploubidis, G.B., 2019. The role of intelligence in decision-making in early adolescence. British Journal of Developmental Psychology 37, 101–111. https://doi.org/10.1111/bjdp.12261
Fornito, A., Bullmore, E.T., 2012. Connectomic Intermediate Phenotypes for Psychiatric Disorders. Front. Psychiatry 3. https://doi.org/10.3389/fpsyt.2012.00032
Fornito, A., Zalesky, A., Bullmore, E., 2016. Fundamentals of Brain Network Analysis. Academic Press.
Fraley, C., Raftery, A.E., 1999. MCLUST: Software for Model-Based Cluster and Discriminant Analysis. Journal of Classification 16, 297–306.
French, B.F., Finch, W.H., 2008. Multigroup Confirmatory Factor Analysis: Locating the Invariant Referent Sets. Structural Equation Modeling: A Multidisciplinary Journal 15, 96–113. https://doi.org/10.1080/10705510701758349
Fried, E.I., 2020. Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. Psychological Inquiry 31, 271–288. https://doi.org/10.1080/1047840X.2020.1853461
Fried, E.I., Cramer, A.O.J., 2017. Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. Perspect Psychol Sci 12, 999–1020. https://doi.org/10.1177/1745691617705892

Fried, E.I., Robinaugh, D.J., 2020. Systems all the way down: embracing complexity in mental health research. BMC Medicine 18, 205. https://doi.org/10.1186/s12916-020-01668-w

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441. https://doi.org/10.1093/biostatistics/kxm045

Friston, K.J., Price, C.J., 2003. Degeneracy and redundancy in cognitive anatomy. Trends in Cognitive Sciences 7, 151–152. https://doi.org/10.1016/S1364-6613(03)00054-8

Frye, R.E., Hasan, K., Malmberg, B., Desouza, L., Swank, P., Smith, K., Landry, S., 2010. Superior longitudinal fasciculus and cognitive dysfunction in adolescents born preterm and at term: Superior Longitudinal Fasciculus and Cognitive Deficits. Developmental Medicine & Child Neurology 52, 760–766. https://doi.org/10.1111/j.1469-8749.2010.03633.x

Fuhrmann, D., Simpson-Kent, I.L., Bathelt, J., Holmes, J., Gathercole, S., Astle, D., Manly, T., Kievit, R., Kievit, R.A., 2020. A Hierarchical Watershed Model of Fluid Intelligence in Childhood and Adolescence. Cereb Cortex 30, 339–352. https://doi.org/10.1093/cercor/bhz091

Gallos, L.K., Makse, H.A., Sigman, M., 2012. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. PNAS 109, 2825–2830. https://doi.org/10.1073/pnas.1106612109

Garcia-Garcia, I., Neseliler, S., Morys, F., Dadar, M., Yau, Y.H.C., Scala, S.G., Zeighami, Y., Sun, N., Collins, D.L., Vainik, U., Dagher, A., 2020. Relationship between impulsivity, uncontrolled eating and body mass index: a hierarchical model. bioRxiv 348821. https://doi.org/10.1101/348821

Garrett, H.E., 1946. A developmental theory of intelligence. The American Psychologist 1, 372–378. http://dx.doi.org.ezp.lib.cam.ac.uk/10.1037/h0056380

Gates, K.M., Henry, T., Steinley, D., Fair, D.A., 2016. A Monte Carlo Evaluation of Weighted Community Detection Algorithms. Front. Neuroinform. 10. https://doi.org/10.3389/fninf.2016.00045

Gathercole, S.E., Durling, E., Evans, M., Jeffcock, S., Stone, S., 2008. Working memory abilities and children's performance in laboratory analogues of classroom activities. Applied Cognitive Psychology 22, 1019–1037. https://doi.org/10.1002/acp.1407

Gathercole, S.E., Service, E., Hitch, G.J., Adams, A.-M., Martin, A.J., 1999. Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. Applied Cognitive Psychology 13, 65–77. https://doi.org/10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O

Gelman, A., Carlin, J., 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspect Psychol Sci 9, 641–651. https://doi.org/10.1177/1745691614551642

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neuroscience 2, 861–863. https://doi.org/10.1038/13158

Gignac, G.E., 2014. Dynamic mutualism versus g factor theory: An empirical test. Intelligence 42, 89–97. https://doi.org/10.1016/j.intell.2013.11.004

Girn, M., Mills, C., Christoff, K., 2019. Linking brain network reconfiguration and intelligence: Are we there yet? Trends in Neuroscience and Education 15, 62–70. https://doi.org/10.1016/j.tine.2019.04.001

Gläscher, J., Rudrauf, D., Colom, R., Paul, L.K., Tranel, D., Damasio, H., Adolphs, R., 2010. Distributed neural system for general intelligence revealed by lesion mapping. PNAS 107, 4705–4709. https://doi.org/10.1073/pnas.0910397107

Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S., 2011. Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. Front. Neuroinform. 5. https://doi.org/10.3389/fninf.2011.00013

Gottfredson, L.S., 1997. Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. Intelligence 24, 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Graham, M.S., Drobnjak, I., Zhang, H., 2016. Realistic simulation of artefacts in diffusion MRI for validating post-processing correction techniques. NeuroImage 125, 1079–1094. https://doi.org/10.1016/j.neuroimage.2015.11.006

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q.K., Yu, A.B., Kahn, A.E., Medaglia, J.D., Vettel, J.M., Miller, M.B., Grafton, S.T., Bassett, D.S., 2015. Controllability of structural brain networks. Nature Communications 6, 8414. https://doi.org/10.1038/ncomms9414

Gur, R.C., Richard, J., Hughett, P., Calkins, M.E., Macy, L., Bilker, W.B., Brensinger, C., Gur, R.E., 2010. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. Journal of Neuroscience Methods 187, 254–262. https://doi.org/10.1016/j.jneumeth.2009.11.017

Haring, L., Mõttus, R., Koch, K., Trei, M., Maron, E., 2015. Factorial validity, measurement equivalence and cognitive performance of the Cambridge Neuropsychological Test Automated Battery (CANTAB) between patients with first-episode psychosis and healthy volunteers. Psychological medicine 45, 1919–1929. https://doi.org/10.1017/S0033291714003018

Hartung, J., Doebler, P., Schroeders, U., Wilhelm, O., 2018. Dedifferentiation and differentiation of intelligence in adults across age and years of education. Intelligence 69, 37–49. https://doi.org/10.1016/j.intell.2018.04.003

Hegelund, E.R., Flensborg-Madsen, T., Dammeyer, J., Mortensen, E.L., 2018. Low IQ as a predictor of unsuccessful educational and occupational achievement: A register-based study of 1,098,742 men in Denmark 1968–2016. Intelligence 71, 46–53. https://doi.org/10.1016/j.intell.2018.10.002

Henrich, J., Heine, S.J., Norenzayan, A., 2010. Most people are not WEIRD. Nature 466, 29–29. https://doi.org/10.1038/466029a

Henry, T.R., Robinaugh, D., Fried, E.I., 2020. On the control of psychological networks. PsyArXiv. https://doi.org/10.31234/osf.io/7vpz2

Heuvel, M.P. van den, Sporns, O., 2011. Rich-Club Organization of the Human Connectome. J. Neurosci. 31, 15775–15786. https://doi.org/10.1523/JNEUROSCI.3539-11.2011

Hilger, K., Ekman, M., Fiebach, C.J., Basten, U., 2017. Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence. Intelligence 60, 10–25. https://doi.org/10.1016/j.intell.2016.11.001

Hilland, E., Landrø, N.I., Kraft, B., Tamnes, C.K., Fried, E.I., Maglanoc, L.A., Jonassen, R., 2020. Exploring the links between specific depression symptoms and brain structure: A network study. Psychiatry and Clinical Neurosciences 74, 220–221. https://doi.org/10.1111/pcn.12969

Holmes, J., Bryant, A., Gathercole, S.E., the CALM Team, 2019. Protocol for a transdiagnostic study of children with problems of attention, learning and memory (CALM). BMC Pediatrics 19. https://doi.org/10.1186/s12887-018-1385-3

Holzinger, K.J., Swineford, F., 1937. The Bi-factor method. Psychometrika 2, 41–54. https://doi.org/10.1007/BF02287965

Horn, J.L., Cattell, R.B., 1967. Age differences in fluid and crystallized intelligence. Acta Psychologica 26, 107–129. https://doi.org/10.1016/0001-6918(67)90011-X

Horn, J.L., Cattell, R.B., 1966. Refinement and test of the theory of fluid and crystallized general intelligences. Journal of Educational Psychology 57, 253–270. https://doi.org/10.1037/h0023816

Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D.S., Calabresi, P.A., Pekar, J.J., van Zijl, P.C.M., Mori, S., 2008. Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. NeuroImage 39, 336–347. https://doi.org/10.1016/j.neuroimage.2007.07.053

Huizinga, M., Dolan, C.V., van der Molen, M.W., 2006. Age-related change in executive function: Developmental trends and a latent variable analysis. Neuropsychologia, Advances in Developmental Cognitive Neuroscience 44, 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Hülür, G., Wilhelm, O., Robitzsch, A., 2011. Intelligence Differentiation in Early Childhood. Journal of Individual Differences 32, 170–179. https://doi.org/10.1027/1614-0001/a000049

Huttenlocher, P.R., 1990. Morphometric study of human cerebral cortex development. Neuropsychologia 28, 517–527. https://doi.org/10.1016/0028-3932(90)90031-I

Iacobucci, D., Posavac, S.S., Kardes, F.R., Schneider, M.J., Popovich, D.L., 2015. The median split: Robust, refined, and revived. Journal of Consumer Psychology 25, 690–704. https://doi.org/10.1016/j.jcps.2015.06.014

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging 27, 685–691. https://doi.org/10.1002/jmri.21049

Jauk, E., Benedek, M., Dunst, B., Neubauer, A.C., 2013. The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. Intelligence 41, 212–221. https://doi.org/10.1016/j.intell.2013.03.003

Johnson, M.H., 2011. Interactive Specialization: A domain-general framework for human functional brain development? Developmental Cognitive Neuroscience 1, 7–21. https://doi.org/10.1016/j.dcn.2010.07.003

Jones, P., 2020. networktools: Tools for Identifying Important Nodes in Networks.

Jones, P.J., Ma, R., McNally, R.J., 2019. Bridge Centrality: A Network Approach to Understanding Comorbidity. Multivariate Behavioral Research 0, 1–15. https://doi.org/10.1080/00273171.2019.1614898

Jöreskog, K.G., 1999. How Large Can a Standardized Coefficient be?

Jöreskog, K.G., Goldberger, A.S., 1975. Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable. Journal of the American Statistical Association 70, 631–639.
Juan-Espinosa, M., García, L.F., Colom, R., Abad, F.J., 2000. Testing the age related differentiation hypothesis through the Wechsler's scales. Personality and Individual Differences 29, 1069–1075. https://doi.org/10.1016/S0191-8869(99)00254-8
Jung, R.E., Haier, R.J., 2007. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. Behavioral and Brain Sciences 30, 135. https://doi.org/10.1017/S0140525X07001185
Kamali, A., Sair, H.I., Radmanesh, A., Hasan, K.M., 2014. Decoding the superior parietal lobule connections of the superior longitudinal fasciculus/arcuate fasciculus in the human brain. Neuroscience 277, 577–583. https://doi.org/10.1016/j.neuroscience.2014.07.035
Kan, K.-J., van der Maas, H.L.J., Levine, S.Z., 2019. Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? Intelligence 73, 52–62. https://doi.org/10.1016/j.intell.2018.12.004
Karwowski, M., Kaufman, J.C., Lebuda, I., Szumski, G., Firkowska-Mankiewicz, A., 2017. Intelligence in childhood and creative achievements in middle-age: The necessary condition approach. Intelligence 64, 36–44. https://doi.org/10.1016/j.intell.2017.07.001
Kaufman, A.S., 1975. Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. Journal of Consulting and Clinical Psychology 43, 135–147.
Khundrakpam, B.S., Lewis, J.D., Reid, A., Karama, S., Zhao, L., Chouinard-Decorte, F., Evans, A.C., 2017. Imaging structural covariance in the development of intelligence. NeuroImage 144, 227–240. https://doi.org/10.1016/j.neuroimage.2016.08.041
Khurana, A., Romer, D., Betancourt, L.M., Hurt, H., 2018. Modeling Trajectories of Sensation Seeking and Impulsivity Dimensions from Early to Late Adolescence: Universal Trends or Distinct Sub-groups? J Youth Adolesc 47, 1992–2005. https://doi.org/10.1007/s10964-018-0891-9
Kievit, R., Frankenhuis, W.E., Waldorp, L., Borsboom, D., 2013. Simpson's paradox in psychological science: a practical guide. Front. Psychol. 4. https://doi.org/10.3389/fpsyg.2013.00513
Kievit, R., Scholte, H.S., Waldorp, L.J., Borsboom, D., 2016. Inter- and intra-individual differences in fluid reasoning show distinct cortical responses (preprint). Neuroscience. https://doi.org/10.1101/039412
Kievit, R.A., 2020. Sensitive periods in cognitive development: a mutualistic perspective. Current Opinion in Behavioral Sciences, Evolution, Development and Genetics *Sensitive and critical periods* 36, 144–149. https://doi.org/10.1016/j.cobeha.2020.10.007
Kievit, R.A., Brandmaier, A.M., Ziegler, G., van Harmelen, A.-L., de Mooij, S.M.M., Moutoussis, M., Goodyer, I.M., Bullmore, E., Jones, P.B., Fonagy, P., Lindenberger, U., Dolan, R.J., 2018a. Developmental cognitive neuroscience using latent change score models: A tutorial and applications. Developmental Cognitive Neuroscience 33, 99–117. https://doi.org/10.1016/j.dcn.2017.11.007
Kievit, R.A., Davis, S.W., Griffiths, J., Correia, M.M., Cam-CAN, Henson, R.N., 2016. A watershed model of individual differences in fluid intelligence. Neuropsychologia 91, 186–198. https://doi.org/10.1016/j.neuropsychologia.2016.08.008

Kievit, R.A., Fuhrmann, D., Borgeest, G.S., Simpson-Kent, I.L., Henson, R.N.A., 2018b. The neural determinants of age-related changes in fluid intelligence: a pre-registered, longitudinal analysis in UK Biobank. Wellcome Open Res 3. https://doi.org/10.12688/wellcomeopenres.14241.2

Kievit, R.A., Hofman, A.D., Nation, K., 2019. Mutualistic Coupling Between Vocabulary and Reasoning in Young Children: A Replication and Extension of the Study by Kievit et al. (2017). Psychol Sci 30, 1245–1252. https://doi.org/10.1177/0956797619841265

Kievit, R.A., Lindenberger, U., Goodyer, I.M., Jones, P.B., Fonagy, P., Bullmore, E.T., Dolan, R.J., 2017. Mutualistic Coupling Between Vocabulary and Reasoning Supports Cognitive Development During Late Adolescence and Early Adulthood. Psychol Sci 28, 1419–1431. https://doi.org/10.1177/0956797617710785

Kievit, R.A., Romeijn, J.-W., Waldorp, L.J., Wicherts, J.M., Scholte, H.S., Borsboom, D., 2011. Modeling Mind and Matter: Reductionism and Psychological Measurement in Cognitive Neuroscience. Psychological Inquiry 22, 139–157. https://doi.org/10.1080/1047840X.2011.567962

Kievit, R.A., Simpson-Kent, I.L., 2021. It's About Time: Towards a Longitudinal Cognitive Neuroscience of Intelligence, in: The Cambridge Handbook of Intelligence and Cognitive Neuroscience. Cambridge University Press.

Kievit, R.A., van Rooijen, H., Wicherts, J.M., Waldorp, L.J., Kan, K.-J., Scholte, H.S., Borsboom, D., 2012. Intelligence and the brain: A model-based approach. Cognitive Neuroscience 3, 89–97. https://doi.org/10.1080/17588928.2011.628383

Kline, R.B., 2015. Principles and Practice of Structural Equation Modeling: Fourth Edition, 4th edition. ed. Guilford Press, New York.

Koechlin, E., Hyafil, A., 2007. Anterior Prefrontal Function and the Limits of Human Decision-Making. Science 318, 594–598. https://doi.org/10.1126/science.1142995

Koenis, M.M.G., Brouwer, R.M., van den Heuvel, M.P., Mandl, R.C.W., van Soelen, I.L.C., Kahn, R.S., Boomsma, D.I., Hulshoff Pol, H.E., 2015. Development of the brain's structural network efficiency in early adolescence: A longitudinal DTI twin study. Hum. Brain Mapp. 36, 4938–4953. https://doi.org/10.1002/hbm.22988

Kovacs, K., Conway, A.R.A., 2016. Process Overlap Theory: A Unified Account of the General Factor of Intelligence. Psychological Inquiry 27, 151–177. https://doi.org/10.1080/1047840X.2016.1153946

Krijnen, W.P., 2004. Positive loadings and factor correlations from positive covariance matrices. Psychometrika 69, 655–660. https://doi.org/10.1007/BF02289861

Krogsrud, S.K., Fjell, A.M., Tamnes, C.K., Grydeland, H., Due-Tønnessen, P., Bjørnerud, A., Sampaio-Baptista, C., Andersson, J., Johansen-Berg, H., Walhovd, K.B., 2018. Development of white matter microstructure in relation to verbal and visuospatial working memory—A longitudinal study. PLOS ONE 13, e0195540. https://doi.org/10.1371/journal.pone.0195540

Laird, A.R., Fox, P.M., Eickhoff, S.B., Turner, J.A., Ray, K.L., McKay, D.R., Glahn, D.C., Beckmann, C.F., Smith, S.M., Fox, P.T., 2011. Behavioral Interpretations of Intrinsic Connectivity Networks. J Cogn Neurosci 23, 4022–4037. https://doi.org/10.1162/jocn_a_00077

Legg, S., Hutter, M., 2007. A Collection of Definitions of Intelligence. arXiv:0706.3639 [cs].

Levenstein, D., Alvarez, V.A., Amarasingham, A., Azab, H., Gerkin, R.C., Hasenstaub, A., Iyer, R., Jolivet, R.B., Marzen, S., Monaco, J.D., Prinz, A.A., Quraishi, S., Santamaria, F., Shivkumar, S., Singh, M.F., Stockton, D.B., Traub, R., Rotstein, H.G., Nadim, F.,

Redish, A.D., 2020. On the role of theory and modeling in neuroscience. arXiv:2003.13825 [q-bio].
Levine, S.Z., Leucht, S., 2016. Identifying a system of predominant negative symptoms: Network analysis of three randomized clinical trials. Schizophrenia Research 178, 17–22. https://doi.org/10.1016/j.schres.2016.09.002
Locurto, C., Scanlon, C., 1998. Individual Differences and a Spatial Learning Factor in Two Strains of Mice (Mus musculus). Journal of Comparative Psychology 112, 344–352.
Luna, B., 2009. Developmental Changes in Cognitive Control through Adolescence. Adv Child Dev Behav 37, 233–278.
Makropoulos, A., Robinson, E.C., Schuh, A., Wright, R., Fitzgibbon, S., Bozek, J., Counsell, S.J., Steinweg, J., Vecchiato, K., Passerat-Palmbach, J., Lenz, G., Mortari, F., Tenev, T., Duff, E.P., Bastiani, M., Cordero-Grande, L., Hughes, E., Tusor, N., Tournier, J.-D., Hutter, J., Price, A.N., Teixeira, R.P.A.G., Murgasova, M., Victor, S., Kelly, C., Rutherford, M.A., Smith, S.M., Edwards, A.D., Hajnal, J.V., Jenkinson, M., Rueckert, D., 2018. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. NeuroImage 173, 88–112. https://doi.org/10.1016/j.neuroimage.2018.01.054
Marek, S., Tervo-Clemmens, B., Calabro, F.J., Montez, D.F., Kay, B.P., Hatoum, A.S., Donohue, M.R., Foran, W., Miller, R.L., Feczko, E., Miranda-Dominguez, O., Graham, A.M., Earl, E.A., Perrone, A.J., Cordova, M., Doyle, O., Moore, L.A., Conan, G., Uriarte, J., Snider, K., Tam, A., Chen, J., Newbold, D.J., Zheng, A., Seider, N.A., Van, A.N., Laumann, T.O., Thompson, W.K., Greene, D.J., Petersen, S.E., Nichols, T.E., Yeo, B.T.T., Barch, D.M., Garavan, H., Luna, B., Fair, D.A., Dosenbach, N.U.F., 2020. Towards Reproducible Brain-Wide Association Studies. bioRxiv 2020.08.21.257758. https://doi.org/10.1101/2020.08.21.257758
Mareva, S., Holmes, J., 2020. Network models of learning and cognition in typical and atypical learners. https://doi.org/10.31219/osf.io/3tn5m
Marois, R., Ivanoff, J., 2005. Capacity limits of information processing in the brain. Trends in Cognitive Sciences 9, 296–305. https://doi.org/10.1016/j.tics.2005.04.010
Marr, D., Poggio, T., 1976. From Understanding Computation to Understanding Neural Circuitry.
Maslowsky, J., Owotomo, O., Huntley, E.D., Keating, D., 2019. Adolescent Risk Behavior: Differentiating Reasoned And Reactive Risk-taking. J Youth Adolescence 48, 243–255. https://doi.org/10.1007/s10964-018-0978-3
McArdle, J.J., 2013. Exploratory Data Mining Using Decision Trees in the Behavioral Sciences [WWW Document]. Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences. https://doi.org/10.4324/9780203403020-10
McArdle, J.J., Hamagami, F., Meredith, W., Bradway, K.P., 2000. Modeling the dynamic hypotheses of Gf–Gc theory using longitudinal life-span data. Learning and Individual Differences 12, 53–79. https://doi.org/10.1016/S1041-6080(00)00036-4
McGrew, K.S., 2009. CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. Intelligence 37, 1–10. https://doi.org/10.1016/j.intell.2008.08.004
Medaglia, J.D., Ramanathan, D.M., Venkatesan, U.M., Hillary, F.G., 2011. The challenge of non-ergodicity in network neuroscience. Network: Computation in Neural Systems 22, 148–153. https://doi.org/10.3109/09638237.2011.639604

Menzies, L., Goddings, A.-L., Whitaker, K.J., Blakemore, S.-J., Viner, R.M., 2015. The effects of puberty on white matter development in boys. Developmental Cognitive Neuroscience, Proceedings from the inaugural Flux Congress; towards an integrative developmental cognitive neuroscience 11, 116–128. https://doi.org/10.1016/j.dcn.2014.10.002

Meunier, D., Lambiotte, R., Bullmore, E.T., 2010. Modular and Hierarchically Modular Organization of Brain Networks. Front. Neurosci. 4. https://doi.org/10.3389/fnins.2010.00200

Molenaar, P.C.M., Campbell, C.G., 2009. The New Person-Specific Paradigm in Psychology. Curr Dir Psychol Sci 18, 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Naglieri, J.A., 1981. Concurrent validity of the revised Peabody Picture Vocabulary Test. Psychology in the Schools 18, 286–289. https://doi.org/10.1002/1520-6807(198107)18:3<286::AID-PITS2310180306>3.0.CO;2-1

Natu, V.S., Gomez, J., Barnett, M., Jeska, B., Kirilina, E., Jaeger, C., Zhen, Z., Cox, S., Weiner, K.S., Weiskopf, N., Grill-Spector, K., 2019. Apparent thinning of human visual cortex during childhood is associated with myelination. PNAS 116, 20750–20759. https://doi.org/10.1073/pnas.1904931116

Navas-Sánchez, F.J., Alemán-Gómez, Y., Sánchez-Gonzalez, J., Guzmán-De-Villoria, J.A., Franco, C., Robles, O., Arango, C., Desco, M., 2014. White matter microstructure correlates of mathematical giftedness and intelligence quotient: White Matter Microstructure. Human Brain Mapping 35, 2619–2631. https://doi.org/10.1002/hbm.22355

Newman, M.E.J., 2006. Modularity and community structure in networks. Proceedings of the National Academy of Sciences 103, 8577–8582. https://doi.org/10.1073/pnas.0601602103

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.-B., 2005. Valid conjunction inference with the minimum statistic. NeuroImage 25, 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005

Nooner, K.B., Colcombe, S.J., Tobe, R.H., Mennes, M., Benedict, M.M., Moreno, A.L., Panek, L.J., Brown, S., Zavitz, S.T., Li, Q., Sikka, S., Gutman, D., Bangaru, S., Schlachter, R.T., Kamiel, S.M., Anwar, A.R., Hinz, C.M., Kaplan, M.S., Rachlin, A.B., Adelsberg, S., Cheung, B., Khanuja, R., Yan, C., Craddock, C.C., Calhoun, V., Courtney, W., King, M., Wood, D., Cox, C.L., Kelly, A.M.C., Di Martino, A., Petkova, E., Reiss, P.T., Duan, N., Thomsen, D., Biswal, B., Coffey, B., Hoptman, M.J., Javitt, D.C., Pomara, N., Sidtis, J.J., Koplewicz, H.S., Castellanos, F.X., Leventhal, B.L., Milham, M.P., 2012. The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. Frontiers in Neuroscience 6. https://doi.org/10.3389/fnins.2012.00152

Oberski, D., 2016. Mixture Models: Latent Profile and Latent Class Analysis, in: Robertson, J., Kaptein, M. (Eds.), Modern Statistical Methods for HCI, Human–Computer Interaction Series. Springer International Publishing, Cham, pp. 275–287. https://doi.org/10.1007/978-3-319-26633-6_12

Oschwald, J., Guye, S., Liem, F., Rast, P., Willis, S., Röcke, C., Jäncke, L., Martin, M., Mérillat, S., 2019. Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. Reviews in the Neurosciences 31, 1–57. https://doi.org/10.1515/revneuro-2018-0096

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H.L., Dardick, W., Tao, S., 2018. A meta-analysis on the relation between reading and working memory. Psychological Bulletin 144, 48–76. https://doi.org/10.1037/bul0000124

Peng, P., Kievit, R.A., 2020. The Development of Academic Achievement and Cognitive Abilities: A Bidirectional Perspective. Child Development Perspectives 14, 15–20. https://doi.org/10.1111/cdep.12352

Peng, P., Wang, T., Wang, C., Lin, X., 2019. A Meta-analysis on the Relation between Fluid Intelligence and Reading/Mathematics: Effects of Tasks, Age, and Social Economics Status. Psychological Bulletin 145, 189–236. https://doi.org/10.1037/bul0000182

Penke, L., Denissen, J.J.A., Miller, G.F., 2007. The evolutionary genetics of personality. European Journal of Personality 21, 549–587. https://doi.org/10.1002/per.629

Peters, B.D., Ikuta, T., DeRosse, P., John, M., Burdick, K.E., Gruner, P., Prendergast, D.M., Szeszko, P.R., Malhotra, A.K., 2014. Age-Related Differences in White Matter Tract Microstructure Are Associated with Cognitive Performance from Childhood to Adulthood. Biological Psychiatry 75, 248–256. https://doi.org/10.1016/j.biopsych.2013.05.020

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J.-B., Vul, E., Yarkoni, T., 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nature Reviews Neuroscience 18, 115–126. https://doi.org/10.1038/nrn.2016.167

Pons, P., Latapy, M., 2005. Computing Communities in Large Networks Using Random Walks, in: Yolum, pInar, Güngör, T., Gürgen, F., Özturan, C. (Eds.), Computer and Information Sciences - ISCIS 2005, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 284–293. https://doi.org/10.1007/11569596_31

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2011. Functional Network Organization of the Human Brain. Neuron 72, 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

Power, J.D., Schlaggar, B.L., Lessov-Schlaggar, C.N., Petersen, S.E., 2013. Evidence for Hubs in Human Functional Brain Networks. Neuron 79, 798–813. https://doi.org/10.1016/j.neuron.2013.07.035

Putnick, D.L., Bornstein, M.H., 2016. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental Review 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Raz, N., Lindenberger, U., 2011. Only Time will Tell: Cross-sectional Studies Offer no Solution to the Age-Brain-Cognition Triangle—Comment on. Psychological bulletin 137, 790. https://doi.org/10.1037/a0024503

Reader, S.M., Hager, Y., Laland, K.N., 2011. The evolution of primate general and cultural intelligence. Phil. Trans. R. Soc. B 366, 1017–1027. https://doi.org/10.1098/rstb.2010.0342

Reckase, M.D., 2009. Multidimensional Item Response Theory. Springer Science & Business Media.

Ritchie, S.J., Quinlan, E.B., Banaschewski, T., Bokde, A.L.W., Desrivières, S., Flor, H., Frouin, V., Garavan, H., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Martinot, M.-L.P., Nees, F., Orfanos, D.P., Paus, T., Poustka, L., Hohmann, S., Millenet, S., Fröhner,

J., Smolka, M.N., Walter, H., Whelan, R., Schumann, G., 2019. Neuroimaging and genetic correlates of cognitive ability and cognitive development in adolescence. https://doi.org/10.31234/osf.io/8pwd6

Roberts, R.P., Hach, S., Tippett, L.J., Addis, D.R., 2016. The Simpson's paradox and fMRI: Similarities and differences between functional connectivity measures derived from within-subject and across-subject correlations. NeuroImage 135, 1–15. https://doi.org/10.1016/j.neuroimage.2016.04.028

Robinaugh, D.J., Hoekstra, R.H.A., Toner, E.R., Borsboom, D., 2019. The network approach to psychopathology: a review of the literature 2008–2018 and an agenda for future research. Psychological Medicine 1–14. https://doi.org/10.1017/S0033291719003404

Rogers, R.D., Owen, A.M., Middleton, H.C., Williams, E.J., Pickard, J.D., Sahakian, B.J., Robbins, T.W., 1999. Choosing between Small, Likely Rewards and Large, Unlikely Rewards Activates Inferior and Orbital Prefrontal Cortex. J. Neurosci. 19, 9029–9038. https://doi.org/10.1523/JNEUROSCI.19-20-09029.1999

Rohrer, J.M., 2018. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. Advances in Methods and Practices in Psychological Science 1, 27–42. https://doi.org/10.1177/2515245917745629

Román, F.J., Abad, F.J., Escorial, S., Burgaleta, M., Martínez, K., Álvarez-Linera, J., Quiroga, M.Á., Karama, S., Haier, R.J., Colom, R., 2014. Reversed hierarchy in the brain for general and specific cognitive abilities: A morphometric analysis. Human Brain Mapping 35, 3805–3818. https://doi.org/10.1002/hbm.22438

Romer, D., Reyna, V.F., Satterthwaite, T.D., 2017. Beyond stereotypes of adolescent risk taking: Placing the adolescent brain in developmental context. Developmental Cognitive Neuroscience 27, 19–34. https://doi.org/10.1016/j.dcn.2017.07.007

Rosseel, Y., 2012. lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software 48. https://doi.org/10.18637/jss.v048.i02

Savi, A.O., Marsman, M., van der Maas, H.L.J., Maris, G.K.J., 2019. The Wiring of Intelligence. Perspect Psychol Sci 14, 1034–1061. https://doi.org/10.1177/1745691619866447

Sawyer, S.M., Azzopardi, P.S., Wickremarathne, D., Patton, G.C., 2018. The age of adolescence. The Lancet Child & Adolescent Health 2, 223–228. https://doi.org/10.1016/S2352-4642(18)30022-1

Schaie, K.W., 1994. The course of adult intellectual development. American Psychologist 49, 304–313. https://doi.org/10.1037//0003-066X.49.4.304

Scharfen, J., Peters, J.M., Holling, H., 2018. Retest effects in cognitive ability tests: A meta-analysis. Intelligence 67, 44–66. https://doi.org/10.1016/j.intell.2018.01.003

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 of Psychological Research 23–74.

Schmank, C.J., Goring, S.A., Kovacs, K., Conway, A.R.A., 2019. Psychometric Network Analysis of the Hungarian WAIS. Journal of Intelligence 7, 21. https://doi.org/10.3390/jintelligence7030021

Schmiedek, F., Lövdén, M., Oertzen, T. von, Lindenberger, U., 2020. Within-person structures of daily cognitive performance differ from between-person structures of cognitive abilities. PeerJ 8, e9290. https://doi.org/10.7717/peerj.9290

Schneider, W.J., McGrew, K.S., 2012. The Cattell-Horn-Carroll model of intelligence, in: Contemporary Intellectual Assessment: Theories, Tests, and Issues, 3rd Ed. The Guilford Press, New York, NY, US, pp. 99–144.

Schneidman, E., Berry, M.J., Segev, R., Bialek, W., 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440, 1007–1012. https://doi.org/10.1038/nature04701

Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A., King, J., 2006. Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. The Journal of Educational Research 99, 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P.J., Dalley, J.W., Flor, H., Gallinat, J., Garavan, H., Heinz, A., Itterman, B., Lathrop, M., Mallik, C., Mann, K., Martinot, J.-L., Paus, T., Poline, J.-B., Robbins, T.W., Rietschel, M., Reed, L., Smolka, M., Spanagel, R., Speiser, C., Stephens, D.N., Ströhle, A., Struve, M., 2010. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry 15, 1128–1139. https://doi.org/10.1038/mp.2010.4

Schwarz, G., 1978. Estimating the Dimension of a Model. The Annals of Statistics 6, 461–464. https://doi.org/10.1214/aos/1176344136

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., Greicius, M.D., 2007. Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. J Neurosci 27, 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Seidlitz, J., Váša, F., Shinn, M., Romero-Garcia, R., Whitaker, K.J., Vértes, P.E., Wagstyl, K., Kirkpatrick Reardon, P., Clasen, L., Liu, S., Messinger, A., Leopold, D.A., Fonagy, P., Dolan, R.J., Jones, P.B., Goodyer, I.M., Raznahan, A., Bullmore, E.T., 2018. Morphometric Similarity Networks Detect Microscale Cortical Organization and Predict Inter-Individual Cognitive Variation. Neuron 97, 231-247.e7. https://doi.org/10.1016/j.neuron.2017.11.039

Shafto, M.A., Tyler, L.K., Dixon, M., Taylor, J.R., Rowe, J.B., Cusack, R., Calder, A.J., Marslen-Wilson, W.D., Duncan, J., Dalgleish, T., Henson, R.N., Brayne, C., Matthews, F.E., Cam-CAN, 2014. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. BMC Neurology 14, 204. https://doi.org/10.1186/s12883-014-0204-1

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., Giedd, J., 2006. Intellectual ability and cortical development in children and adolescents. Nature 440, 676–679. https://doi.org/10.1038/nature04513

Siegenfeld, A.F., Bar-Yam, Y., 2020. An Introduction to Complex Systems Science and Its Applications [WWW Document]. Complexity. https://doi.org/10.1155/2020/6105872

Silva, R., Gramacy, R.B., 2010. Gaussian Process Structural Equation Models with Latent Variables. arXiv:1002.4802 [cs, stat].

Simmonds, D.J., Hallquist, M.N., Asato, M., Luna, B., 2014. Developmental stages and sex differences of white matter and behavioral development through adolescence: A longitudinal diffusion tensor imaging (DTI) study. NeuroImage 92, 356–368. https://doi.org/10.1016/j.neuroimage.2013.12.044

Smith, S.M., 2002. Fast robust automated brain extraction. Human Brain Mapping 17, 143–155. https://doi.org/10.1002/hbm.10062
Solé-Casals, J., Serra-Grabulosa, J.M., Romero-Garcia, R., Vilaseca, G., Adan, A., Vilaró, N., Bargalló, N., Bullmore, E.T., 2019. Structural brain network of gifted children has a more integrated and versatile topology. Brain Struct Funct 224, 2373–2383. https://doi.org/10.1007/s00429-019-01914-9
Spearman, C., 1927. The abilities of man. Macmillan, New York, NY, US.
Spearman, C., 1904. "General Intelligence," Objectively Determined and Measured. The American Journal of Psychology 15, 201–292.
Sporns, O., Betzel, R.F., 2016. Modular Brain Networks. Annu Rev Psychol 67, 613–640. https://doi.org/10.1146/annurev-psych-122414-033634
Sporns, O., Honey, C.J., Kötter, R., 2007. Identification and Classification of Hubs in Brain Networks. PLOS ONE 2, e1049. https://doi.org/10.1371/journal.pone.0001049
Steinberg, L., 2008. A social neuroscience perspective on adolescent risk-taking. Developmental Review, Current Directions in Risk and Decision Making 28, 78–106. https://doi.org/10.1016/j.dr.2007.08.002
Steinberg, L., 2005. Cognitive and affective development in adolescence. Trends in Cognitive Sciences 9, 69–74. https://doi.org/10.1016/j.tics.2004.12.005
Tamnes, C.K., Herting, M.M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R.E., Güroğlu, B., Raznahan, A., Sowell, E.R., Crone, E.A., Mills, K.L., 2017. Development of the Cerebral Cortex across Adolescence: A Multisample Study of Inter-Related Longitudinal Changes in Cortical Volume, Surface Area, and Thickness. J. Neurosci. 37, 3402–3412. https://doi.org/10.1523/JNEUROSCI.3302-16.2017
Tamnes, C.K., Østby, Y., Walhovd, K.B., Westlye, L.T., Due-Tønnessen, P., Fjell, A.M., 2010. Intellectual abilities and white matter microstructure in development: A diffusion tensor imaging study. Human Brain Mapping 31, 1609–1625. https://doi.org/10.1002/hbm.20962
Thomson, G., 1947. Charles Spearman, 1863-1945. Obituary Notices of Fellows of the Royal Society 5, 373–385. https://doi.org/10.1098/rsbm.1947.0006
Thomson, G.H., 1939. The factorial analysis of human ability. Houghton Mifflin, Oxford, England.
Thomson, G.H., 1919. The Hierarchy of Abilities. British Journal of Psychology, 1904-1920 9, 337–344. https://doi.org/10.1111/j.2044-8295.1919.tb00231.x
Thomson, G.H., 1916. A hierarchy without a general factor. British Journal of Psychology, 1904-1920 8, 271–281. https://doi.org/10.1111/j.2044-8295.1916.tb00133.x
Tideman, E., Gustafsson, J.-E., 2004. Age-related differentiation of cognitive abilities in ages 3–7. Personality and Individual Differences 36, 1965–1974. https://doi.org/10.1016/j.paid.2003.09.004
Tombu, M.N., Asplund, C.L., Dux, P.E., Godwin, D., Martin, J.W., Marois, R., 2011. A Unified attentional bottleneck in the human brain. PNAS 108, 13426–13431. https://doi.org/10.1073/pnas.1103583108
Tu, Y.-K., Gunnell, D., Gilthorpe, M.S., 2008. Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon – the reversal paradox. Emerg Themes Epidemiol 5, 2. https://doi.org/10.1186/1742-7622-5-2
Uddin, L.Q., Yeo, B.T.T., Spreng, R.N., 2019. Towards a Universal Taxonomy of Macro-scale Functional Human Brain Networks. Brain Topogr 32, 926–942. https://doi.org/10.1007/s10548-019-00744-6

Urger, S.E., De Bellis, M.D., Hooper, S.R., Woolley, D.P., Chen, S.D., Provenzale, J., 2015. The Superior Longitudinal Fasciculus in Typically Developing Children and Adolescents: Diffusion Tensor Imaging and Neuropsychological Correlates. Journal of Child Neurology 30, 9–20. https://doi.org/10.1177/0883073813520503

van den Heuvel, M.P., Bullmore, E.T., Sporns, O., 2016. Comparative Connectomics. Trends in Cognitive Sciences 20, 345–361. https://doi.org/10.1016/j.tics.2016.03.001

van den Heuvel, M.P., Sporns, O., 2013. Network hubs in the human brain. Trends in Cognitive Sciences, Special Issue: The Connectome 17, 683–696. https://doi.org/10.1016/j.tics.2013.09.012

van der Maas, H., Savi, A.O., Hofman, A., Kan, K.J., Marsman, M., 2021. The Network Approach to General Intelligence. https://doi.org/10.31234/osf.io/e96m8

van der Maas, H.L.J., Dolan, C.V., Grasman, R.P.P.P., Wicherts, J.M., Huizenga, H.M., Raijmakers, M.E.J., 2006. A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. Psychological Review 113, 842–861. https://doi.org/10.1037/0033-295X.113.4.842

van der Maas, H.L.J., Kan, K.-J., Marsman, M., Stevenson, C.E., 2017. Network Models for Cognitive Development and Intelligence. Journal of Intelligence 5, 16. https://doi.org/10.3390/jintelligence5020016

Vertes, P.E., Alexander-Bloch, A.F., Gogtay, N., Giedd, J.N., Rapoport, J.L., Bullmore, E.T., 2012. Simple models of human brain functional networks. Proceedings of the National Academy of Sciences 109, 5868–5873. https://doi.org/10.1073/pnas.1111738109

Vértes, P.E., Bullmore, E.T., 2015. Annual Research Review: Growth connectomics – the organization and reorganization of brain networks during normal and abnormal development. Journal of Child Psychology and Psychiatry 56, 299–320. https://doi.org/10.1111/jcpp.12365

Vijayakumar, N., Allen, N.B., Youssef, G., Dennison, M., Yücel, M., Simmons, J.G., Whittle, S., 2016. Brain development during adolescence: A mixed-longitudinal investigation of cortical thickness, surface area, and volume. Human Brain Mapping 37, 2027–2038. https://doi.org/10.1002/hbm.23154

Vijayakumar, N., Mills, K.L., Alexander-Bloch, A., Tamnes, C.K., Whittle, S., 2018. Structural brain development: A review of methodological approaches and best practices. Developmental Cognitive Neuroscience, Methodological Challenges in Developmental Neuroimaging: Contemporary Approaches and Solutions 33, 129–148. https://doi.org/10.1016/j.dcn.2017.11.008

Vincent, J.L., Kahn, I., Snyder, A.Z., Raichle, M.E., Buckner, R.L., 2008. Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. Journal of Neurophysiology 100, 3328–3342. https://doi.org/10.1152/jn.90355.2008

Voelkle, M.C., Brose, A., Schmiedek, F., Lindenberger, U., 2014. Toward a Unified Framework for the Study of Between-Person and Within-Person Structures: Building a Bridge Between Two Research Paradigms. Multivariate Behavioral Research 49, 193–213. https://doi.org/10.1080/00273171.2014.889593

Volkow, N.D., Koob, G.F., Croyle, R.T., Bianchi, D.W., Gordon, J.A., Koroshetz, W.J., Pérez-Stable, E.J., Riley, W.T., Bloch, M.H., Conway, K., Deeds, B.G., Dowling, G.J., Grant, S., Howlett, K.D., Matochik, J.A., Morgan, G.D., Murray, M.M., Noronha, A., Spong, C.Y., Wargo, E.M., Warren, K.R., Weiss, S.R.B., 2018. The conception of the ABCD

study: From substance use to a broad NIH collaboration. Developmental Cognitive Neuroscience 32, 4–7. https://doi.org/10.1016/j.dcn.2017.10.002
Vollmar, C., O'Muircheartaigh, J., Barker, G.J., Symms, M.R., Thompson, P., Kumari, V., Duncan, J.S., Richardson, M.P., Koepp, M.J., 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. https://doi.org/10.1016/j.neuroimage.2010.03.046
Vollmer, B., Lundequist, A., Mårtensson, G., Nagy, Z., Lagercrantz, H., Smedler, A.-C., Forssberg, H., 2017. Correlation between white matter microstructure and executive functions suggests early developmental influence on long fibre tracts in preterm born adolescents. PLOS ONE 12, e0178893. https://doi.org/10.1371/journal.pone.0178893
Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. Perspect Psychol Sci 4, 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x
Walhovd, K.B., Fjell, A.M., Westerhausen, R., Nyberg, L., Ebmeier, K.P., Lindenberger, U., Bartrés-Faz, D., Baaré, W.F.C., Siebner, H.R., Henson, R., Drevon, C.A., Knudsen, G.P.S., Ljøsne, I.B., Penninx, B.W.J.H., Ghisletta, P., Rogeberg, O., Tyler, L., Bertram, L., Consortium, L., 2018. Healthy minds 0–100 years: Optimising the use of European brain imaging cohorts ("Lifebrain"). European Psychiatry 50, 47–56. https://doi.org/10.1016/j.eurpsy.2017.12.006
Wandell, B.A., 2016. Clarifying Human White Matter. Annual Review of Neuroscience 39, 103–128. https://doi.org/10.1146/annurev-neuro-070815-013815
Warne, R.T., Burningham, C., 2019. Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. Psychological Bulletin 145, 237–272. https://doi.org/10.1037/bul0000184
Watkins, M.W., Beaujean, A.A., 2014. Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition. School Psychology Quarterly 29, 52–63. https://doi.org/10.1037/spq0000038
Wechsler, D., 2011. Wechsler Abbreviated Scales of Intelligence-Second Edition (WASI-II).
Wechsler, D., 2008. Wechsler adult intelligence Scale–Fourth edition (WAIS–IV).
Wechsler, D., 2005. Wechsler Individual Achievement Test-Second UK Edition (WIAT-II).
Wechsler, D., 1999. Wechsler Abbreviated Scales of Intelligence.
Wendelken, C., Ferrer, E., Ghetti, S., Bailey, S.K., Cutting, L., Bunge, S.A., 2017. Frontoparietal Structural Connectivity in Childhood Predicts Development of Functional Connectivity and Reasoning Ability: A Large-Scale Longitudinal Investigation. The Journal of Neuroscience 37, 8549–8558. https://doi.org/10.1523/JNEUROSCI.3726-16.2017
Woolgar, A., Duncan, J., Manes, F., Fedorenko, E., 2018. Fluid intelligence is supported by the multiple-demand system not the language system. Nature Human Behaviour 2, 200–204. https://doi.org/10.1038/s41562-017-0282-3
Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., Roca, M., Antoun, N., Manes, F., Duncan, J., 2010. Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. PNAS. https://doi.org/10.1073/pnas.1007928107
Woolley, A.W., Aggarwal, I., Malone, T.W., 2015. Collective Intelligence and Group Performance. Curr Dir Psychol Sci 24, 420–424. https://doi.org/10.1177/0963721415599543

Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., Malone, T.W., 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. Science 330, 686–688. https://doi.org/10.1126/science.1193147

Yin, W., Li, T., Hung, S.-C., Zhang, H., Wang, L., Shen, D., Zhu, H., Mucha, P.J., Cohen, J.R., Lin, W., 2020. The emergence of a functionally flexible brain during early infancy. PNAS 117, 23904–23913. https://doi.org/10.1073/pnas.2002645117

Zadelaar, J.N., Weeda, W.D., Waldorp, L.J., Van Duijvenvoorde, A.C.K., Blankenstein, N.E., Huizenga, H.M., 2019. Are individual differences quantitative or qualitative? An integrated behavioral and fMRI MIMIC approach. NeuroImage 202, 116058. https://doi.org/10.1016/j.neuroimage.2019.116058

Zhang, S., Malatesha Joshi, R., 2020. Longitudinal relations between verbal working memory and reading in students from diverse linguistic backgrounds. Journal of Experimental Child Psychology 190, 104727. https://doi.org/10.1016/j.jecp.2019.104727

Zmigrod, L., Rentfrow, P.J., Robbins, T.W., 2018. Cognitive underpinnings of nationalistic ideology in the context of Brexit. PNAS 115, E4532–E4540. https://doi.org/10.1073/pnas.1708960115

Citation Diversity Statement

Recent work has demonstrated that gender[4,5] (as well as racial/ethnic[6]) minorities are systemically under-cited in neuroscience. I estimated the gender citation practices of this dissertation using the *Journal of Cognitive Neuroscience*'s Gender Citation Balance Index tool (GCBI-alyzer, https://postlab.psych.wisc.edu/gcbialyzer/). Overall, 212 DOIs were successfully categorised and metrics for man (first author)/man (last author), woman (first author)/man (last author), man (first author)/woman (last author), and woman (first author)/woman (last author) were calculated. Percentages for the 212 DOIs were as follows: MM = 56.1%, WM = 17%, MW = 12.7%, and WW = 14.2%. The GCBIs were as follows: MM = 0.376, WM = -0.493, MW = 0.179, and WW = -0.05. This indicates that I over-cited MM and MW (> 0 indicates over-citation) and under-cited WM and WW (< 0 denotes under-citation). Note that these estimates assume a binary paradigm of gender (man or woman) and, therefore, do not account for non-binary identifications of gender. Furthermore, these estimates are based on the authorship practices of one journal, *Journal of Cognitive Neuroscience*. Even though this thesis can be classified as (developmental) cognitive neuroscience, citations also include purely behavioural (e.g., cognitive psychology) and neuroscience (e.g., physical properties of the brain such as small-worldness) literature.

Nevertheless, the estimated metrics indicate that I have under-cited literature involving women and disproportionately over-cited research involving men. Lastly, although I will not add references merely to ensure a desirable proportion, I will work towards a more balanced citation rate in future research. This will include expanding my literature search to a wider net of journals, to seek out relevant literature.

---

[4] Dworkin, Jordan D., et al. "The extent and drivers of gender imbalance in neuroscience reference lists." *Nature neuroscience* 23.8 (2020): 918-926.

[5] Fulvio, Jacqueline M., Ileri Akinnola, and Bradley R. Postle. "Gender (im) balance in citation practices in cognitive neuroscience." *Journal of Cognitive Neuroscience* 33.1 (2021): 3-7.

[6] Bertolero, Maxwell A., et al. "Racial and ethnic imbalance in neuroscience reference lists and intersections with gender." *BioRxiv* (2020).